Genet. Sel. Evol. 37 (2005) 291–313

DOI: 10.1051/gse:2005003

Original article

Genomic contributions in livestock gene introgression programmes

Eileen WALL[a,b*], Peter M. VISSCHER[a], Frédéric HOSPITAL[c], John A. WOOLLIAMS[b]

[a] Institute of Cell, Animal and Population Biology, Ashworth Laboratories, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JG, Scotland
[b] Roslin Institute, Roslin, Midlothian, EH25 9PS, Scotland
[c] Station de génétique végétale, INRA/UPS/INA-PG, Ferme du Moulon, 91190 Gif sur Yvette, France

(Received 6 August 2004; accepted 19 December 2004)

Abstract – The composition of the genome after introgression of a marker gene from a donor to a recipient breed was studied using analytical and simulation methods. Theoretical predictions of proportional genomic contributions, including donor linkage drag, from ancestors used at each generation of crossing after an introgression programme agreed closely with simulated results. The obligate drag, the donor genome surrounding the target locus that cannot be removed by subsequent selection, was also studied. It was shown that the number of backcross generations and the length of the chromosome affected proportional genomic contributions to the carrier chromosomes. Population structure had no significant effect on ancestral contributions and linkage drag but it did have an effect on the obligate drag whereby larger offspring groups resulted in smaller obligate drag. The implications for an introgression programme of the number of backcross generations, the population structure and the carrier chromosome length are discussed. The equations derived describing contributions to the genome from individuals from a given generation provide a framework to predict the genomic composition of a population after the introgression of a favourable donor allele. These ancestral contributions can be assigned a value and therefore allow the prediction of genetic lag.

introgression / genomic contributions / linkage drag / backcross / genetic lag

1. INTRODUCTION

There is a wealth of genetic diversity among breeds and lines of livestock and it is reasonable to assume that some alleles have become fixed in populations before artificial selection was introduced. Commercial lines or breeds are

* Corresponding author: eileen.wall@sac.ac.uk

unlikely to contain all the best alleles for traits considered of economic importance (*e.g.*, [24]). Developments in molecular genetics have led to the uncovering of individual alleles or regions of the genome that have an effect on traits of interest that may wish to be utilised in commercial livestock lines; *e.g.*, the halothane sensitivity locus [13], the RN gene [19] and the estrogen receptor locus [21] in pigs; the double muscling gene [9] and polled gene [1] in cattle; and callipyge gene in sheep [6] (increasing meat yield), in pigs (meat quality), in cattle (welfare) in pigs (reproductive). These loci cover traits of relevance to welfare, health, fitness, quality, productive and reproductive performance. A relevant point of interest in the latter example is that the beneficial allele increasing fecundity is found at much higher frequency in a non-commercial line of pigs (Meishan).

Gene introgression can be used as a tool for genetic improvement by the introduction of new alleles into a population to address challenges facing current breeding goals [18]. Having detected an allele of interest from a non-commercial (donor) line, the aim of introgression is to fix that allele into a commercial (recipient) population whilst minimising the contribution of the donor genome, thereby minimising the loss of beneficial alleles from the commercial population. Introgression involves (i) a number of generations of backcrossing of individuals carrying the desired allele to the recipient breed to obtain further heterozygotes that have an increasingly higher proportion of the recipient breed genome; followed by (ii) an *inter se* cross among those heterozygotes to breed individuals that are homozygous for the desired allele (*e.g.*, [11, 15, 18, 26]). This technique has been made considerably more attractive with the advent of DNA markers to track the alleles that derive from donor and recipient breeds.

In the course of gene introgression many donor alleles linked to the desired allele are incorporated into the genome of the recipient line by a phenomenon called linkage drag [2]. Linkage drag is defined as the length of donor genome segment surrounding a gene of introgression. The linkage drag segment is important as it may incorporate other less favourable alleles and drag them into the commercial population and the risk of this is related to its length. In addition to a donor genome segment around the gene of introgression there may be other residual donor segments, both on the chromosome of introgression and on other chromosomes. Several authors (*e.g.*, [14, 20, 23]) have examined the prediction of the expected length of the linkage drag in backcross breeding programmes. Stam and Zeven [23] showed that the proportion of donor genome, both around the introgressed gene and elsewhere on the chromosome, can be large, *e.g.*, 32 cM, for a 100 cM chromosome after six generations of

backcrossing. This theoretical work has been verified by practical examples, not only in plants in wheat [28] and barley [3] but in sheep [28] the latter being a rare livestock example. Whilst experimental observations give a general validation of the accuracy of the theoretical prediction they are open to errors depending on the extent of DNA information and the size of the studies. The impact of the linkage drag after the *inter se* cross of an introgression programme has not been addressed.

Marker information associated with the desired allele (*e.g.*, flanking markers) and markers specific to the recipient line can be used with marker assisted selection (MAS) protocols to minimise the donor contamination in the genome during backcross phase [5, 10, 18]. This possibility is most powerful in plant populations where the offspring group size is large, offering considerable selection opportunity and successive backcross generations can be carried out over relatively short periods of time. The study of Hospital *et al.* [18] showed that 98.5% of the recipient genome can be recovered in four generations of backcrossing when using MAS to speed the recovery of recipient genome on non-carrier chromosomes during introgression (compared to six generations without using MAS). The study of Frisch and Melchinger [7] and Hospital [16] also examined the use of MAS to reduce the length of the linkage drag. The overall impact of the introgression programme including donor contribution and loss of selection opportunity in the recipient breed can be measured as a genetic lag, *i.e.*, the difference between the non-introgressed commercial population and the offspring of the *inter se* cross. Gama *et al.* [8] and Visscher and Haley [25] developed equations to describe genetic lag between an introgression population and a commercial population considering only the number of backcross generations. However they did not consider the impact of linkage drag on genetic lag.

There are important differences between plant and livestock introgression programmes and these revolve around family size and generation interval.

(i) The selection intensities during a livestock introgression programme are lower than those achieved in plant breeding and this severely limits the selection of favourable recombinants at flanking markers, particularly if alleles at multiple loci are to be introgressed. These limits arise from the practical constraints on the size of the introgression population and/or the biological constraints on offspring group size.

(ii) Plant breeders would tend to select a single individual with the most favourable set of recombinants to continue using for backcrossing. Livestock introgression programmes will wish to continue with multiple carriers and to finish the introgression programme with a viable breeding population. Threats

to viability arise from genetic bottlenecks in both the donor contributions and in the *inter se* cross (which will form a new inter breed population). Considerations of the parental and offspring numbers change the statistical properties of some of the parameters (*e.g.*, the obligate drag: the component of the donor genome that cannot be removed by selection).

(iii) The recovery of the recipient breed genome and/or the reduction in the linkage drag length does not remove all ancestral genome that will contribute to genetic lag. This is of particular importance to livestock introgression programmes as the generation interval is much larger than that of plants. The commercial populations will make significant rates of genetic change in the time it takes for introgression. Having predictions of the recipient genomic contributions from different generations of the introgression scheme is therefore important to allow a breeder to optimise the design of the programme.

The purpose of the present study was to investigate the genomic state of the carrier and non-carrier chromosomes after introgression of a desired allele through to the *inter se* cross through analytical methods and simulation studies. The components of genetic lag, including the predictions of the recipient genomic contributions, were described. The parameters considered are (i) the number of backcrosses, (ii) the length of the carrier chromosome and total genome, and (iii) the structure of the populations, *i.e.*, numbers of parents and offspring per generation on the predicted genomic contributions are studied. The results derived are validated by simulation studies.

2. MATERIALS AND METHODS

2.1. Populations, structures and notation

The introgression of a marker for a desired allele at a target locus, with the proximal end of the chromosome s M distant from the target locus, is performed by crossing donor and recipient individuals to create F_1 individuals born at time 0, followed by T generations of backcrossing (creating individuals born at times 1 to T), and an *inter se* cross born at time $T + 1$. D refers to individuals of the donor breed used to initiate the introgression programme and R refers to individuals of the recipient breed used as parents at each generation of backcrossing. BC refers to the backcross heterozygous offspring, subsequently used as parents, and IC refers to the offspring of the *inter se* cross that are homozygous for the desired allele. This is summarised in Table I and Figure 1.

Subscript t is used to denote time, so R_t refers to the group of recipient breed parents used to produce BC_t offspring at time t ($t = 1$ to T). The special case,

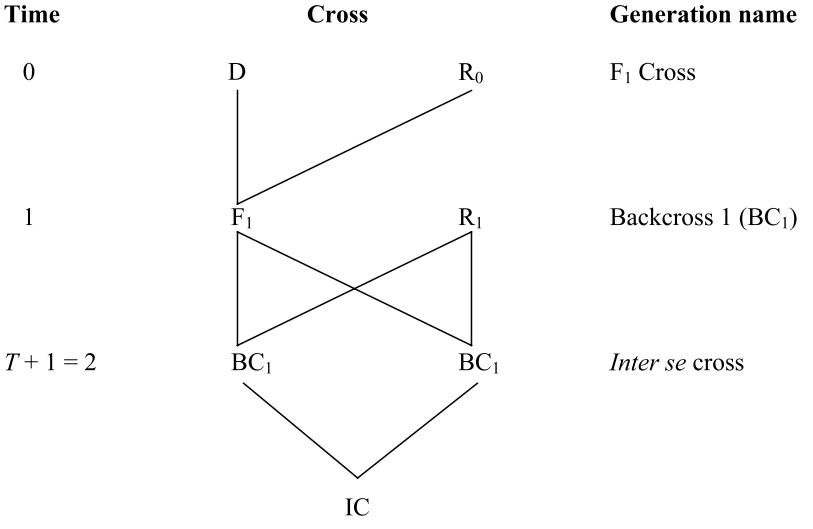

Figure 1. Diagram of an introgression scheme with one backcross generation showing population groups, generation names and time period.

Table I. Design of an introgression of a gene from a donor breed (D) into a recipient population (R) with an F_1, T generations of backcrossing (BC) followed by an *inter se* cross (IC), with a description of the ancestral origins of the alleles.

Gen^n	Cross	Offspg	Description	Contributions
0	$R_0 \times D$	F_1	Recipients crossed with F_1 donors to create	D, R_0
1	$R_1 \times F_1$	BC_1	F_1 individuals backcrossed to recipients	D, R_0, R_1
2	$R_2 \times BC_1$	BC_2	BC_1 individuals backcrossed to recipients	D, R_0, R_1, R_2
.....	.....	.....	.....	.....
T	$R_T \times BC_{T-1}$	BC_T	BC_{T-1} individuals backcrossed to recipients	$D, R_0, R_1, \ldots R_{T-1}, R_T$
$T + 1$	$BC_T \times BC_T$	IC	BC_T individuals crossed for the *inter se* cross	$D, R_0, R_1, \ldots R_{T-1}, R_T$

R_0 is used to denote parents of the recipient breed used to produce the F_1 cross. In the introgression programme the number of mating pairs at each backcross generation is N with n offspring per mating (*e.g.*, $N\,D \times N\,R_0$ produce $Nn\ F_1$ offspring). All carrier individuals at the end of backcrossing are used for the *inter se* cross. The lengths of the carrier chromosome and total genome are assumed to be l and L Morgan, respectively. The total length of all non-carrier chromosomes in the genome is therefore given by $(L - l)$.

$\pi(G)$ is the proportion of total alleles in *IC* from population group G, where G can be the donor (D), recipient groups R_0 to R_T or offspring group F_1, BC_1 to BC_T. The subscripts C or NC denote carrier or non-carrier chromosomes respectively. To develop predictions, the total genomic contributions of population groups will be described by summation after considering separately the proximal and distal contributions. A list of notation is given in the Appendix.

2.2. Theoretical considerations on genomic contributions

Haldane's mapping function [12] is used which assumes no interference in crossing over events. It is also assumed that loci are uniformly distributed over the chromosome map, (*i.e.*, segments of equal length will contain an equal number of loci). To extrapolate the genomic contributions to *IC* individuals it is necessary to note that the *inter se* cross may be treated as one additional generation of backcrossing since recombinations in the parents are with the recipient genome (this is not so for any further generations of *inter se* crossing).

2.2.1. Carrier chromosome

For each generation, we consider the chromosome from the previous *BC* generation post-recombination, *i.e.* for a BC_t generation, we consider the chromosome inherited from the BC_{t-1} parent, not the parent from the recurrent breed. This also applies for the *IC* generation, since for a single *IC* generation each chromosome can be treated separately.

Consider a target locus at position s on a chromosome of total length l (all distances in Morgans, no interference in recombination is assumed). Consider a locus X on the carrier chromosome, located at distance x from the target locus, corresponding to recombination rate r between X and the target locus:

$$r = r[x] = 1/2(1 - e^{-2x}). \tag{1}$$

The conditional probability of the genotype at locus X is computed for chromosomes carrying the donor allele at the target locus (*i.e.*, after selection). In the *IC* generation the chromosome was inherited from the BC_T parent, and there was either a recombination between X and the target locus, or not. The allele at the target locus was always inherited from the previous *BC* parent (*i.e.*, BC_{T-1}), because of selection. If there was no recombination, then the allele at locus X also comes from the BC_{T-1} parent. If there was a recombination, then

the allele at X comes from the R_T parent. Hence, in the IC generation, given that the target locus carries the donor allele, the allele at locus X was inherited from R_T with a probability r, and was inherited from BC_{T-1} with a probability of $(1 - r)$. Extending the argument back a further generation, the probability that R_{T-1} parent transmitted an allele to the IC population is $(1 - r)r$, and that for a BC_{T-2} parent is $(1 - r)^2$.

Extending this, given that the target locus carries the donor allele, the probability that the allele at locus X in generation IC was inherited from the R_{T-k} parent is:

$$f[R_{T-k},\ x] = (1 - r[x])^k r[x]. \tag{2}$$

Conversely, the conditional probability that the allele at locus X in generation IC was inherited from the donor (D) parent is:

$$f[D, x] = (1 - r[x])^{T+1}. \tag{3}$$

The contribution $\pi(G)$ of each ancestral group G on the whole chromosome is then simply obtained by integrating along the chromosome on each side of the target locus:

$$\pi(G) = \int_{x=0}^{s} f[G, x]\mathrm{d}x + \int_{x=0}^{l-s} f[G, x]\,\mathrm{d}x \tag{4}$$

where density f is taken from equation (2) for recipient groups R_0 to R_T ($k = 0$ to $k = T$), or from equation (3) for donor D, and where $r[x]$ is taken from equation (1).

The linkage drag, in this study, is defined strictly as the intact segment around the target locus that originates from the donor line. The expectation of the length of the linkage drag segment after an F_1 and t generations of backcrossing for a locus in position s is termed $\delta(s, t)$. Therefore, for the IC group the linkage drag is given by

$$\delta_{IC} = \delta(s, T + 1).$$

Following Hanson [14], the linkage drag proximal to the target locus at position s is given by:

$$\delta(s, t) = t^{-1}\left(1 - \mathrm{e}^{-ts}\right). \tag{5}$$

Analogously the linkage drag distal to the target locus is given by $t^{-1}(1 - \mathrm{e}^{-t(l-s)})$.

In addition to that segment there may be other contributions from the donor genome to the chromosome since the total proportion of the carrier chromosome that is inherited from the donor line must be greater than, or equal to,

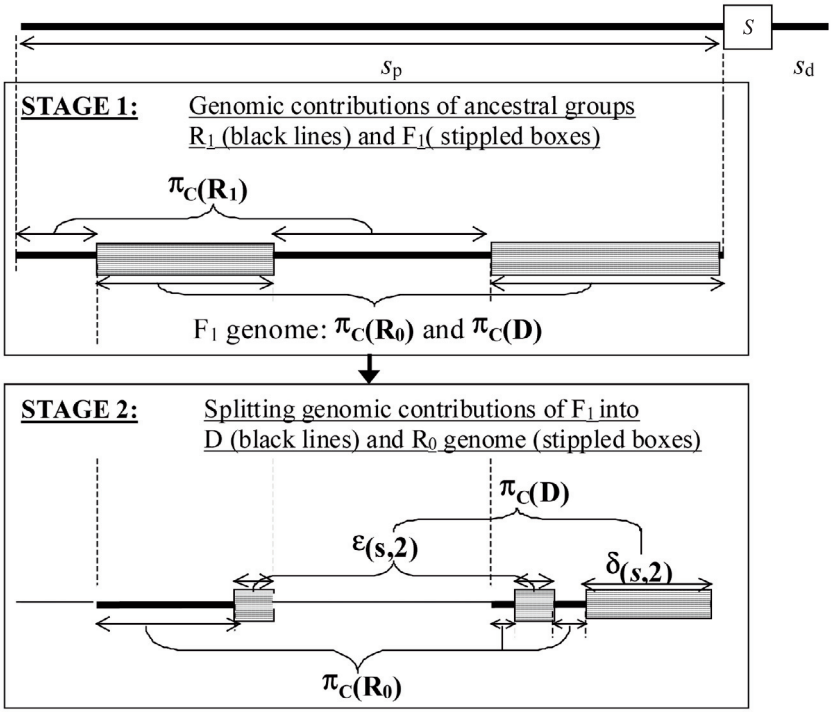

Figure 2. Diagram illustrating the derivation of ancestral contributions proximal to a target locus at position *s*.

the contribution from linkage drag. The proportion of residual donor genome outside the linkage drag in *IC* after *T* generations of backcrossing (see Fig. 2) can be given by:

$$\varepsilon_{IC} = \pi_C(D) - \delta_{IC}. \tag{6}$$

2.2.2. Non-carrier chromosomes

The length of donor genome on the non-carrier chromosomes, *i.e.*, $\pi_{NC}(D)$ is $1/2^{T+1}(L-l)$. The contribution of $R_0, \ldots R_T$ to the non-carrier chromosomes follows the same pattern and are:

$$\begin{cases} \pi_{NC}(R_T) = 1/2(L-l) \\ \pi_{NC}(R_{T-1}) = 1/2^2(L-l) \\ \pi_{NC}(R_{T-t}) = 1/2^{t+1}(L-l) \\ \pi_{NC}(R_0) = \pi_{NC}(D) = 1/2^{T+1}(L-l). \end{cases} \tag{7}$$

2.3. Genetic lag

Summing the preceding equations (6) and (7) gives the entire genomic contributions of ancestral groups post-introgression.

$$\pi_E(G) = \pi_{NC}(G) + \pi_C(G). \tag{8}$$

Each $\pi_E(G)$ can be weighted for the genetic worth of individuals at any given point in time assuming, for example, an infinitesimal genetic model. Assume M_D gives the difference in background genetic merit between recipient (R_0) and donor (D) genome and the commercial population has a genetic gain of ΔG per generation. Then the genetic lag, ΔI, may be estimated as:

$$\Delta I = (T+1)\Delta G - \left[\pi_E(D)M_D + \sum_{t=0}^{T+1} \pi_E(R_t)(T+1-t)\Delta G\right]. \tag{9}$$

The first term represents genetic gain in the commercial population and the second is the total gain for commercial traits during introgression. This assumes the scheme is unable to make any selection other than for the donor target allele, which is likely to be the situation for ruminants, but less likely for pigs and poultry where the greater reproductive rate may allow for some concurrent selection.

2.4. Variance of linkage drag and obligate drag

Stam and Zeven [23] derived the variance of the donor genomic contributions $\pi_C(D)$ to the carrier chromosome with the same assumptions and same probability density as used by Hanson [14], derived as $\delta(s, t)$ in equation (5). We checked the accuracy of these predictions *via* simulation studies.

In addition we have examined the obligate drag, ω_{IC}, defined as that component of the donor genome that cannot be removed by selection after the *IC* generation (see Fig. 3). Obligate drag is the minimum length of the donor segment around the introgressed allele that is shared identical-by-descent by all carier chromosomes in the IC population. Obligate drag can be predicted from the known distribution of segment length [23] and from order statistics [4] assuming independence between families and across generations. The distribution of segment length x from a marker to one side of the chromosome after t generations of backcrossing is, $f(x) = t^{-tx}$. The probability density function of the shortest segment length x_1 from the marker to one side of the chromosome from a sample of k, assuming a chromosome with infinite length,

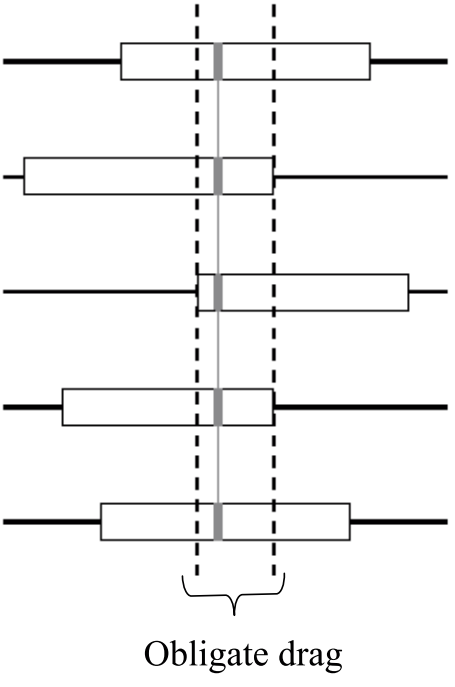

Figure 3. Diagram illustrating the obligate drag on a set of carrier chromosomes. A clear box denotes the linkage drag segment, the target locus is denoted in grey and dotted back lines denote the obligated drag.

is, $f(x_1) = kt^{-ktx_1}$ [4] with expectation $E(x_1) = 1/(kt)$. In our case, the sample size is N (k), because N carriers are selected for breeding each generation, and we consider the segment lengths on both sides of the marker. Therefore, the prediction of obligate drag, in cM, when $l = 1$ M is $\omega_{IC} = 200/(Nt)$.

2.5. Simulations

The initial cross for the introgression scheme was assumed to be between two divergent lines that are fixed for alternative alleles at each locus. The carrier chromosome was simulated using crossing-over events, occurring as a Poisson process, which were generated assuming Haldane's mapping function without interference. The N parents for the next generation were selected at random among the offspring heterozygous for the target locus. At the end of the backcross phase all heterozygous individuals were used as parents of the *inter se* cross. Only the offspring homozygous at the target locus for the donor allele were considered in summaries of the *IC* population.

The results examine the validity of the predictions of linkage drag and recipient genomic contributions. The following were derived from the genome

of each *IC* individual: (i) the proportional genomic contribution from each of the ancestral population groups, $\pi_C(D)$, $\pi_C(R_t)$, $\pi_{NC}(D)$, $\pi_{NC}(R_t)$ and (ii) the total length of the segment of intact donor genome, δ_{IC}, on each side of s. From these the expected linkage drag, its variance and the obligate drag were calculated.

The population parameters used were: parental population size (N = 10, 20, 50); offspring group size (n = 2, 3, 4, 5); the length of the carrier chromosome (l = 0.5, 1, 2, 4, 8 M) and the total number of backcross generations (T = 3, 6, 10, 20). The location of the target locus $s$was varied and the major results are derived for $s = l/2$ and s = 0.1 M. The simulations were run for 500 replicates for each set of parameters studied.

3. RESULTS

3.1. Validity of the theoretical predictions

3.1.1. Linkage drag and proportional genomic contributions

Theoretical predictions of the proportional contributions of donor and recipient genome to the carrier chromosomes of *IC*, *i.e.* $\pi_C(D)$, $\pi_C(R_t)$, and the linkage drag δ_{IC} with its standard deviation within replicate, are compared to simulation results in Table II (for N = 20, n = 2, l = 1 M, T = 6 and s = 0.1 M, *i.e.*, the target locus is towards one end). The predictions of proportional contributions in other schemes are shown in Figure 4 and compared to simulation results. The predictions for all schemes studied are very accurate. Given the derivations above, note that the observed precision for all values of T demonstrates that prediction for the linkage drag and contributions for all intermediate generations within a particular scheme will be accurate.

The predictions of intact donor genome in *IC* around the target locus using equation (5) after T generations of backcrossing were accurate for a target locus in all positions. Predictions of δ_{IC} using Hanson [14], which assumes a central position for the target locus, were overestimates when this assumption was broken. Predictions of $\pi_C(D)$ by Stam and Zeven [23], which assume a random location for the target locus, were also overestimates unless the target locus was centrally positioned. For example, for the parameters N = 20, n = 2, l = 1 M, T = 6 and s = 0.1 M, the simulations and predictions derived in this paper gave δ_{IC} = 0.21 and $\pi_C(D)$ = 0.24. However, the prediction of δ_{IC} by Hanson [14] and $\pi_C(D)$ by Stam and Zeven [23] were 0.28 and 0.27, respectively.

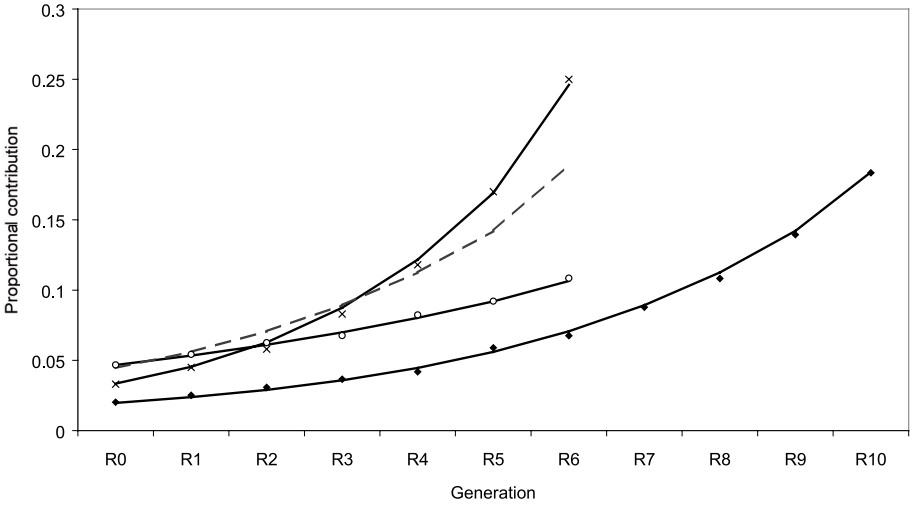

Figure 4. Comparison of simulated and predicted results $\pi_C(G)$, where G, is an ancestral population, expressed as a proportion of l for a variety of schemes. Symbols represent simulations and lines represent predictions. (♦) $T = 10$, $l = 1$, $s = 0.5$ M; (○) $T = 6$, $l = 0.5$, $s = 0.25$ M and (×) $T = 6$, $l = 1$, $s = 0.1$ M. The broken line indicates predictions for (×) when $s = 0.5$ M.

Table II. Comparison of the simulated and predicted genomic contributions from ancestral population groups to the carrier chromosome in *IC*.*

		Simulation	Prediction
	$\pi_C(D)$	0.24 ± 0.003	0.243
	$\pi_C(R_0)$	0.033 ± 0.002	0.034
	$\pi_C(R_1)$	0.045 ± 0.002	0.045
Ancestral	$\pi_C(R_2)$	0.058 ± 0.002	0.061
group	$\pi_C(R_3)$	0.083 ± 0.002	0.084
	$\pi_C(R_4)$	0.118 ± 0.002	0.118
	$\pi_C(R_5)$	0.170 ± 0.002	0.169
	$\pi_C(R_6)$	0.250 ± 0.002	0.246
Linkage drag		0.220 ± 0.004	0.215
s.d. of linkage drag		0.129 ± 0.002	0.114
Obligate drag		0.017 ± 0.001	0.011

* $N = 20$, $n = 2$, $l = 1$ M, $T = 6$ and $s = 0.1$ M.

Figure 4 shows the proportional contributions from the recipient ancestral groups $\pi_C(R_T)$ changed significantly when the target locus is non-centrally placed compared to when it is centrally placed. These changes were correctly predicted using equation (4).

Table III. Comparison of simulated and predicted results for the obligate drag length expressed in cM (±s.e.) over a number of backcross generations and population structures when $l = 1$ M and $s = 0.5$ M.

Population structure	Population group	Simulation	Prediction
$N = 20$, $n = 2$	BC_1	10.5 ± 0.75	10
	BC_2	5.5 ± 0.40	5
	BC_3	5.4 ± 0.38	3.3
	BC_4	4.4 ± 0.36	2.5
	BC_5	3.9 ± 0.28	2
	BC_6	2.3 ± 0.24	1.7
$N = 20$, $n = 4$	BC_1	9.1 ± 0.55	10
	BC_2	5.9 ± 0.45	5
	BC_3	4.3 ± 0.27	3.3
	BC_4	3.8 ± 0.26	2.5
	BC_5	3.5 ± 0.21	2
	BC_6	2.4 ± 0.14	1.7

3.1.2. Obligate drag

It can be seen in Table III that the prediction of obligate drag, ω_{IC}, is close to the simulation result when $n = 2$ and 4 but only in early backcross generations. As a general result varying N to 10 or 50 made no difference to this result. In summary, the assumptions underlying the prediction for obligate drag begin to break down for $t > 3$ leading to significant over-predictions.

The simulation results showed that when $n = 4$ the obligate drag was, on average, smaller than when $n = 2$. When $n = 4$ there are more carriers to select N from for the next population and by chance some of the candidates may have a small obligate drag. However, when $n = 2$ on average only N carriers are produced within a generation and therefore all must go forward to the next generation, so there is no chance to select a carrier with a smaller obligate drag. In fact, population size decreases slightly below N when $n = 2$.

3.2. Effects of carrier chromosome length and position of target locus on proportional genomic contributions

The predictions were used to explore the impact of carrier chromosome length ($l = 0.5$, 1, 2, 4, 8 M, $T = 6$) and position of target locus on the genomic contributions of ancestral populations. Figure 5 shows that as l increases the $\pi_C(D)$ decreases. This decline is mirrored by the early backcross generations and compensated for by increases in total contributions from

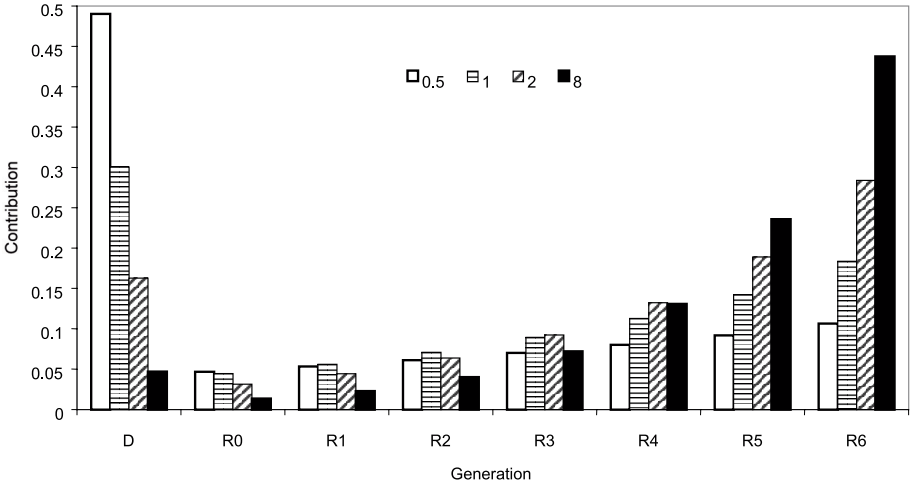

Figure 5. Proportional genomic contributions from ancestral population groups to the carrier chromosome of *IC* as a function of *l* M, where *l* = 0.5, 1, 2 and 8 M.

later R_t generations. For example, when s is centrally placed and $l = 0.5$ M, $\pi_C(D) \sim 50\%$ and $\pi_C(R_6) \sim 10\%$, but when $l = 8$ M $\pi_C(D) \sim 5\%$ and $\pi_C(R_6) \sim 44\%$.

Whilst the proportional linkage drag length changes dramatically for the different chromosome lengths studied the absolute length is relatively constant. With this example ($T = 6$, $N = 20$, $n = 2$) the linkage drag varies over a very narrow range around 28 cM, except for when $l = 0.5$ when linkage drag is 24 cM.

As the location of the target locus on the carrier chromosome approaches the chromosome end the $\pi_C(D)$ decreases, primarily due to a decrease in linkage drag (Fig. 6), and is compensated for by an increase in $\pi_C(R_t)$. On larger chromosomes, δ_{IC} remains relatively constant, similar to a centrally-placed target, until the target locus approached the very edge of the chromosome. In this case the amount of donor genome decreased dramatically. For example, Figure 6 shows that when $T = 6$, only when $s = 8$ cM or less does the linkage drag differ markedly for a centrally placed target (approximately 20 cM compared to 28 cM) and this is relatively insensitive to l.

3.3. Effect of number of backcross generations on proportional genomic contributions

The proportional donor contributions to the *IC* cross, $\pi_C(D)$, are highest when the number of backcross generations (T) is small (Fig. 7, with $l = 1$ M

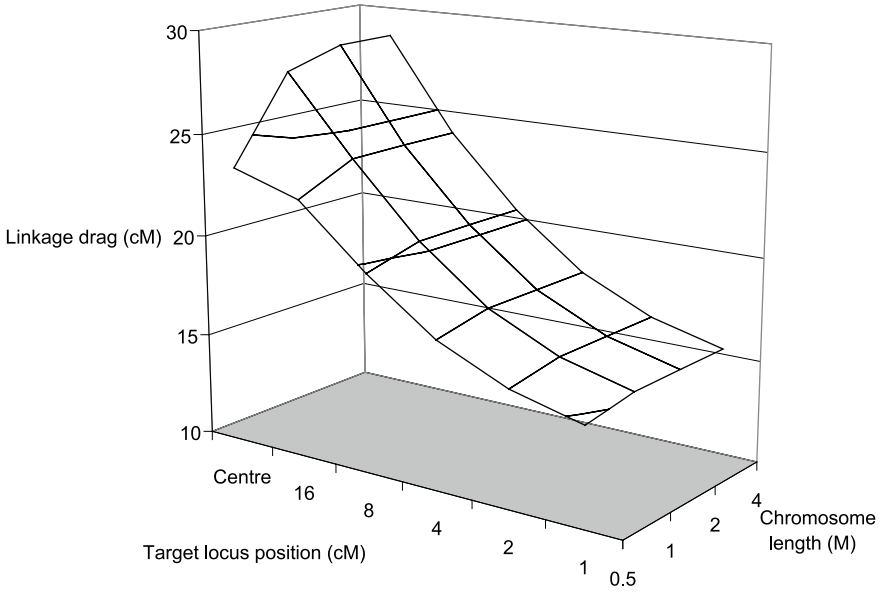

Figure 6. Length of linkage drag in *IC* as a function of carrier chromosome length (l M) and target locus position (s M) after six backcross generations.

and $s = 0.5$ M). Analytical results using equation (4) show that $\pi_C(R_T)$ is independent of T, and this is supported by the simulations. For constant T, s and l, recipient contributions are a function of the number of backcross generations that have been performed since the introduction of the said ancestral recipient group, and δ_{IC} declined with T for all s, in agreement with the predictions given by Hanson [14] for the special case of $s = {}^1/_2 l$.

3.4. Prediction of genetic lag

Equation (9) can be used to estimate the genetic lag of an introgression programme in comparison to a commercial population undergoing selection for commercial traits. A hypothetical example applicable to a livestock production system would be the introgression of the Belgian Blue double muscling allele into the commercial dairy Holstein-Friesian breed. The double muscling gene is located on chromosome 2 [9]. The following assumptions and parameters were used: (i) an initial breed difference, $M_D = 5500$ litres of milk/lactation (7000 *vs.* 1500 litres), (ii) genetic improvement, $\Delta G = 105$ litres, or approximately 1.5% of the mean production, per annum, (iii) cows in the introgression programme are mated to the top available bulls at each backcross generation,

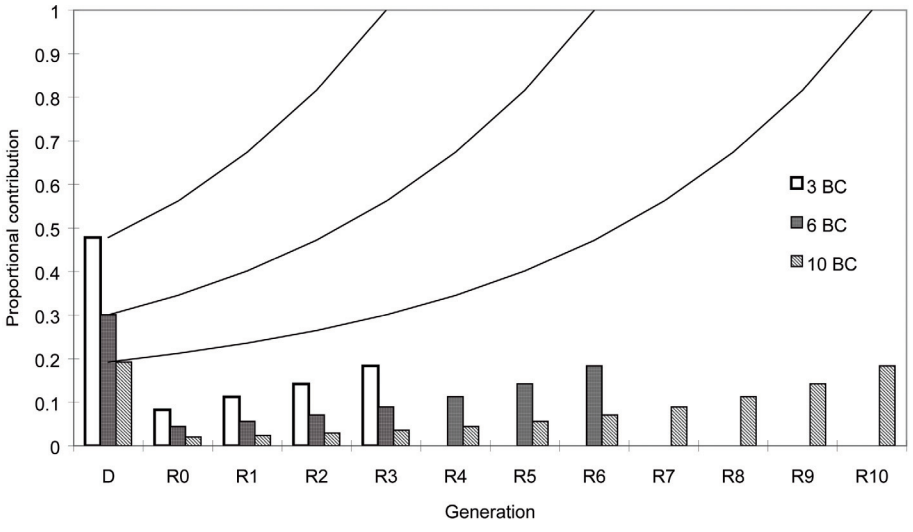

Figure 7. Proportional genomic contributions from ancestral population groups to the carrier chromosome of IC as a function of T. Lines indicate the cumulative values of genomic contributions over the generations.

(iv) a generation interval of 4 years, (v) $L \approx 35$ M for the cattle genome, (vi) chromosome 2, $l \approx 1.25$ M and (vii) $s = 0.1$ M for the double muscling locus. The example will not account for the substitution effect of the target allele, but only examines the genetic lag for the commercial traits present in the recipient line before introgression.

For an introgression programme incorporating $T = 4$ backcross generations (20 years) the genetic lag was predicted to be ~1516 litres (9428 *vs.* 7912) or just under 12 years of selection. This can be compared to the prediction of Visscher and Haley [25], which ignored the genomic contributions of the carrier chromosome and underestimated this value by 8%. As the position of the target locus becomes more central the linkage drag becomes larger and the fractional error between the two predictions increases slightly (increasing to 10.5% when $s = 62.5$ cM).

In an introgression programme with only a few (2 or 3) generations of backcrossing the donor genome makes up a large proportion of both the carrier and non-carrier chromosomes alike, and therefore the effect of the carrier chromosome is not as important. However if the number of backcrosses increases $\pi_{NC}(D)$ decreases rapidly, but δ_{IC} and $\pi_C(D)$ do not decrease as quickly and so become a more important source of genetic lag.

4. DISCUSSION

Using analytical methods and simulation studies the genomic contributions of individuals used at each generation to the genome of individuals after a gene introgression programme were quantified. The prediction of donor and recipient individuals genomic contributions agreed closely with the simulated results for all population structures (N and n), number of backcross generations (T) and length of carrier chromosome (l) studied thereby validating the predictions derived by this study. This study has shown that carrier chromosome length influenced the proportional linkage drag. These influences were quantified with a high degree of precision for all ancestral group contributions and a prediction of genetic lag was developed from them.

The linkage drag and predictions of the linkage drag and donor contributions have been well studied by many authors (*e.g.*, [7, 14, 20, 23]). Hanson [14] defined the linkage drag as the length of intact donor genome segments either side of a target locus and derived its distribution for a centrally placed locus. This distribution was used to predict linkage drag for any target locus position and showed that the assumption of a central position is an upper bound to the size of the linkage drag in all cases. The importance of the linkage drag is that undesirable alleles associated with the donor breed located within this region may be "dragged" into the recipient breed (*e.g.*, [29]). The drag is a function of T and l and, since the length of the carrier chromosome cannot be changed, the one method to control this risk is the number of backcross generations. The introgression programme can therefore be designed so that the expected length of the linkage drag segment does not contain any known deleterious or undesirable donor alleles. If suitable flanking markers are available, another method to control this risk (not examined here) is the selection of favourable recombinants minimising the expected length and variation of the linkage drag segment [16–18].

We have studied the obligate drag, which is the portion of the donor genome on the carrier chromosome in IC that would never be removed by selection for recombinants. Predictions for the obligate drag can be obtained from the same density as the expected drag [14], but these were found to be reliable only when $t < 3$. With two offspring per mating the parental pair, on average, replaces itself with 1 carrier offspring to represent the family line in the next generation. In this situation there is no variation within a family in terms of the linkage drag and obligate drag and the only variation across the population lies between family lines due to recombination. However a higher offspring group size introduces within family variation to the scheme and results potentially

in a wider distribution of linkage drag resulting in lower obligate drag lengths than predicted, as shown in Table III.

The obligate drag is the portion of the donor genome on the carrier chromosome in IC that would never be removed by selection for recombinants. Simple predictions for the obligate drag agreed when t was low but diverged from simulation results in later backcross generations. The use of order statistics assumes independence between generations. This is not the case with this study as all carrier offspring of all N carriers have an equal chance of being selected as one of N carriers for the next generation. This means that some families may contribute more carriers to the next generation at the expense of another family and therefore there is dependence within the population and between generations.

This study is assumes that the location of the donor gene/QTL is known with full precision and the marker for selection is in the gene/QTL. Previous studies have focussed on methodologies that use flanking markers to select carriers of the desired gene. The use of flanking markers in introgression will mean that the drag (linkage and obligate) will be even longer than stated in this study because of the inaccurate estimates of gene position.

As with the linkage drag the number of backcross generations and the length of the carrier chromosome also affect the recipient contributions to the carrier chromosome. Figures 5 and 7 show how the genomic contribution from individuals used in the F_1 cross to the carrier chromosome decreases as T and l increases. These trends are vital in the design of livestock introgression schemes as the proportion of recipient genome from earlier backcross generations also adds to the genetic lag. These predicted contributions from the different ancestral population groups could be used for improving the predictions of genetic lag and other parameters such as identity by descent. Using background selection can increase the recovery of the recipient genome either by minimising the linkage drag using flanking markers or selecting for recipient alleles on non-carrier chromosomes [7, 16, 17, 26]. These equations give an accurate prediction of the recipient contributions to the carrier chromosome, $\pi_C(R)$. This could help to optimise the type of background selection carried out during introgression. In some previous studies on introgression more than one *inter se* cross is considered to maximise the number of homozygous (for the target locus) individuals post introgression. In animal studies time may not permit a second or third *inter se* cross. As animals are not mated to recipient population during an *inter se* cross no new genome will be introduced to the system and therefore the introgression population will lag even further behind the commercial population.

Population parameters may not be as easy to vary in some species as described in this paper and this is particularly so in livestock. The obligate drag was shown to decrease slightly for larger offspring group sizes ($n = 4$) which is good for livestock species with large offspring group size such as pigs or chickens. The higher litter sizes allows for the potential selection among carrier offspring for individuals with smaller linkage drag and obligate drag segments in these species. If the population were made up of single offspring bearing animals (cattle, sheep, etc.) an introgression programme would be difficult to maintain and selection pressure for recipient breed traits low after selection of mating pairs with the desired genotypes. A multiple ovulation and embryo transfer (MOET) scheme may be useful in increasing the numbers of progeny per female in these situations, reducing genetic lag and allowing scope for selection of individuals to reduce donor contamination or allow for concurrent selection on recipient traits or genotypes. Utilising marker and reproductive technology means that ruminant introgression schemes could be successful, potentially resulting in a viable breeding population carrying the target donor allele and the favourable commercial traits of the recipient line.

The prediction of genetic lag presented in equation (9) includes the prediction of linkage drag and recipient individuals' genomic contributions to the carrier chromosome in the formula. Gama *et al.* [8] and Visscher and Haley [25] ignored the effect of genomic contributions of the carrier chromosome and therefore their prediction underestimates genetic lag. The difference between the two predictions is largest when the proportional length of the carrier chromosome in the genome is high and using the method of Visscher and Haley [25] would lead to a highly inaccurate prediction of potential lag. For example, the chicken genome is made up of 6 pairs of macrochromosomes and 30 pairs of microchromosomes [22], for example Chromosome 1 is 3.8 M and approximately 17% of the genome. Underestimating the genetic lag may effect a breeder's decision on the type of programme to use (*e.g.*, the minimum number of backcross generations needed to achieve a certain acceptable genetic lag given the proportional length of the carrier chromosome) and may make the difference between success and failure of the programme and commercial viability.

The prediction of genetic lag does not include selection in the introgression programme for commercial traits or the economic improvement due to the extra value earned from new donor trait included. However, whilst incomplete, equation (9) serves as a basis for estimating the monetary cost of an introgression programme. Modifications would need to include the premium attached to a new commercial product as well as the cost of the programme in

a cost-benefit analysis in different populations to pin-point the best and most cost effective type of introgression programme for different populations.

The results suggest that introducing the double-muscling gene into dairy cattle, but such a scheme does not appear to be viable (large genetic lag, unfeasibility of active selection, etc.). However, other examples of introgression in livestock species have shown less extreme losses post introgression. For example, the study of Wall *et al.* [27] investigated the effect of population size and number of backcross generations on genetic lag and linkage drag around a target region reducing back fat and increasing fecundity found in the Chinese Meishan breed when backcrossed to a commercial Large White population. The results show genetic lag reduces in early backcross generations but after five generations the lag approached an asymptote to 5% difference between the introgression and commercial populations. Although there was genetic lag for commercial traits (decrease average daily gain costing £5/pig), the beneficial effect of an extra piglet per littler (£10/piglet) and the back fat allele compensate for this lag.

Despite conflicting public opinion on the introduction of novel genes into commercial plant and animal populations, the traditional means of introgression of new-to-the-breed genes into many breeds may become more desirable with DNA technology (*e.g.*, to help combat diseases endemic to some livestock populations such as scrapie). These new-to-the-breed genes in commercial lines will allow the breeder to meet the challenges faced by commercial livestock industries. Introgression is more likely to be commercially acceptable in comparison to the creation of transgenic animals as a method of incorporating new genes into livestock populations because it does not involve genetic engineering. Traditional introgression allows scope for more selection in the future as the genetic variance may be maintained with careful planning. This may not be the case for some of the genetically engineered options. The methods outlined in this study will give simple means of planning such introgression programmes to help control the genetic uncertainties involved.

ACKNOWLEDGEMENTS

EEW was funded by a Walsh Fellowship (Teagasc, Ireland), the Irish Cattle Breeding Federation and a Postgraduate Studentship from the British Department for the Environment, Food and Rural Affairs (DEFRA). JAW is grateful to DEFRA for financial support. We thank Ian White for helpful discussions.

REFERENCES

[1] Brenneman R.A., Davis S.K., Sanders J.O., Burns B.M., Wheeler T.C., Turner J.W., Taylor J.F., The polled locus maps to BTA1 in a *Bos indicus* × *Bos taurus* cross, J. Hered. 87 (1996) 156–161.

[2] Brinkman M.A., Frey K.J., Yield component analysis of oat isolines that produce different grain yields, Crop Sci. 17 (1977) 165–168.

[3] Brown A.H.D., Lawrence G.J., Jenkin M., Douglass J., Gregory E., Linkage drag in backcross breeding in barley, J. Hered. 80 (1989) 234–239.

[4] Cox D.R., Hinkley D.V., Appendix 2, Theoretical statistics, Chapman and Hall, London, 1974.

[5] Franklin I.R., Improving the efficiency of backcrossing programs using DNA markers, Proc. Assoc. Advmt. Anim. Breed. Genet. 13 (1999) 357–360.

[6] Freking B.A., Murphy S.K., Wylie A.A., Rhodes S.J., Keele J.W., Leymaster K.A., Jirtle R.L., Smith T.P.L., Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals, Genome Res. 12 (2002) 1496–1506.

[7] Frisch M., Melchinger A.E., The length of intact donor chromosome segment around a target gene in marker-assisted backcrossing, Genetics 157 (2001) 1343–1356.

[8] Gama L.T., Smith C., Gibson J.P., Transgene effects, introgression strategies and testing schemes in pigs, Anim. Prod. 54 (1992) 427–440.

[9] Grobet L., Martin L.J.R., Poncelet D., Pirottin D., Brouwers B., Riquet J., Schoeberlin A., Dunner S., Ménissier F., Massabanda J., Fries R., Hanset R., Georges M., A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle, Nat. Genet. 17 (1997) 71–74.

[10] Groen A.F., Smith C., A stochastic simulation study of the efficiency of marker-assisted introgression in livestock, J. Anim. Breed. Genet. 112 (1995) 161–170.

[11] Groen A.F., Timmermans M.M.J., The use of genetic markers to increase the efficiency of introgression – a simulation study, Proc. 19 Worlds Poultr. C. 1 (1992) 523–527.

[12] Haldane J.B.S., The combination of linkage values and the calculation of distances between the loci of linked factors, J. Genet. 8 (1919) 299–309.

[13] Hanset R., Dasnois C., Scalais S., Michaux C., Grobet L., Introgression into the pietrain genome of the normal allele at the locus for halothane sensitivity, Genet. Sel. Evol. 27 (1995) 77–88.

[14] Hanson W.D., Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing or selfing, Genetics 44 (1959) 833–837.

[15] Hillel J., Schaap T., Haberfield A., Jeffreys A.J., Plotzky Y., Cahaner A., Lavi U., DNA fingerprints applied to gene introgression in breeding programs, Genetics 124 (1990) 783–789.

[16] Hospital F., Size of donor chromosome segments surrounding introgressions and the reduction of linkage drag in marker-assisted backcross programs, Genetics 158 (2001) 1363–1379.

[17] Hospital F., Charcosset A., Marker-assisted introgression of quantitative trait loci, Genetics 147 (1997) 1469–1485.
[18] Hospital F., Chevalet C., Mulsant P., Using markers in gene introgression breeding programs, Genetics 132 (1992) 1199–1210.
[19] Looft C., Milan D., Jeon J.T., Paul S., Reinsch N., Rogel-Gaillard C., Rey V., Amarger V., Robic A., Kalm E., Chardon P., Andersson L., A high-density linkage map of the RN region in pigs, Genet. Sel. Evol. 32 (2000) 321–329.
[20] Naveira H., Barbadilla A., The theoretical distribution of lengths of intact chromosome segments around a locus held heterozygous with backcrossing in a diploid species, Genetics 130 (1992) 205–209.
[21] Rothschild M., Jacobson C., Vaske D., Tuggle C., Wang L.Z., Short T., Eckardt G., Sasaki S., Vincent A., McLaren D., Southwood O., vanderSteen H., Mileham A., Plastow G., The estrogen receptor locus is associated with a major gene influencing litter size in pigs, Proc. Natl. Acad. Sci. USA 93 (1996) 201–205.
[22] Smith J., Bruley C.K., Paton I.R., Dunn I., Jones C.T., Windsor D., Morrice D.R., Law A.S., Masabanda J., Sazanov A., Waddington D., Fries R., Burt D.W., Differences in gene density on chicken macrochromosomes and microchromosomes, Anim. Genet. 31 (2000) 96–103.
[23] Stam P., Zeven A.C., The theoretical proportion of the donor genome in near isogenic lines of self-fertilizers bred by backcrossing, Euphytica 30 (1981) 227–238.
[24] Tanksley S.D., McCouch S.R., Seed banks and molecular maps: Unlocking the genetic potential from the wild, Science 277 (1997) 1063–1066.
[25] Visscher P.M., Haley C.S., On the efficiency of marker-assisted introgression, Anim. Sci. 68 (1999) 59–68.
[26] Visscher P.M., Haley C.S., Thompson R., Marker-assisted introgression in backcross breeding programs, Genetics 144 (1996) 1923–1932.
[27] Wall E.E., Woolliams J.A., Visscher P.M., Genetic lag in a Meishan × Large White pig backcross population: A simulation study, Proc. BSAS (2002) 59.
[28] Walling G.A., Dodds K.G., Galloway S.M., Beattie A.E., Lumsden J.M., Lord E.A., Montgomery G.W., McEwan J.C., The consequences of introducing the Booroola Fecundity (*FecB)* gene on liveweight, Proc. BSAS (2000) 43.
[29] Zeven A.C., Knott D.R., Johnson R., Investigation of linkage drag in near isogenic lines of wheat be testing for seedling reaction to races of stem rust, leaf rust and yellow rust, Euphytica 32 (1983) 319–327.

APPENDIX: SUMMARY OF MAIN SYMBOLS AND NOTATION USED IN THE TEXT

General terms	
T	Number of backcross generations in the introgression
L, l	Length of total genome, and length of carrier chromosome
N	Number of mating pairs at each backcross generation
n	Number of offspring per mating
s	Proximal distance to the end of the carrier chromosome for the target locus
t	Time variable
$\pi(G)$	Contribution of population G to the genome of IC
$\delta(s,t)$	Linkage drag for position s after t backcrosses
δ_{IC}, ω_{IC}	Linkage drag and obligate drag in the genome of the IC
Population group terms	
F_1	Offspring of the initial cross between the donor and recipient breed
BC	Backcross heterozygous offspring
IC	*Inter se* cross offspring homozygous for the target allele
D	Donor individuals
R	Recipient individuals used in F_1 cross (R_0) and each backcross (R_1, R_T)
Subscript terms	
NC	Denotes a non-carrier chromosome
C	Denotes a carrier chromosome
E	Entire genome (a weighted aggregate of NC and C)
Genetic lag terms	
M_D	Difference in background genotype between recipient and donor populations
ΔG	Genetic gain per generation in the commercial population
ΔI	Genetic lag in introgression population for commercial traits