Genotype Selection to Rapidly Breed Congenic Strains

Michael M. Weil,* Barry W. Brown† and Dan M. Serachitopol†

**Department of Experimental Radiation Oncology and †Department of Biomathematics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030*

Manuscript received January 6, 1997
Accepted for publication March 21, 1997

ABSTRACT

Congenic strains can now be constructed guided by the transmission of DNA markers. This allows not only selection for transmission of a desired, donor-derived differential region but also selection against the transmission of unwanted donor origin genomic material. The additional selection capacity should allow congenic strains to be produced in fewer generations than is possible with random backcrosses. Here, we consider modifications of a standard backcross breeding scheme to produce congenic mice by the inclusion of genotype-based selective breeding strategies. Simulation is used to evaluate the consequences of each strategy on the number of chromosomes that contain unwanted, donor-derived genetic material and the average length of this unwanted donor DNA for each backcross generation. Our prototypic strategy was to choose a single mouse to sire each generation using criteria designed to select against the transmission of chromosomes, other than the one containing the replacement genomic region, that contain any donor origin sequence at all. This chromosome elimination strategy resulted in an average of 16.4 chromosomes free of donor DNA in mice of the third backcross (N_3) generation. A strategy based solely on positive selection for the replacement region required six backcross generations to achieve the same results.

THE advent of dense genetic maps has made the mapping of loci involved in multigene traits in the mouse practicable (DIETRICH *et al.* 1992). Customarily, crosses are set up between two strains that differ for the trait under investigation. Their F_2 progeny are phenotyped for the trait and genotyped using genetic markers, generally amplification polymorphisms based on simple sequence repeats. Several computer programs such as Mapmaker QTL and Mapmanager QT allow the data to be scanned for regions of the genome that are associated with the trait (PATERSON *et al.* 1988; LINCOLN *et al.* 1992) (K. F. MANLEY, Map Manager home page, http://mcbio.med.buffalo.edu/mapmgr.html). Molecular cloning of a gene involved in a multigene trait requires refinement of the rough localization provided by this initial screen. This is best accomplished by the construction of congenic strains that contain the locus of interest in the differential genomic region, thus removing the effects of other loci.

A congenic strain is derived from an established inbred strain through the replacement of a selected genomic region, the differential region, with the equivalent region from another strain. This replacement can be accomplished by repeated backcrosses with selection for the new genomic region followed by intercrosses. The strain that provides the bulk of the background genome is referred to as the inbred partner and the strain that provides the differential region of the genome is the donor strain. Until recently, the differential region was selected indirectly through selection for a phenotypic trait encoded within that region (*i.e.*, cell surface antigen, coat color, protein allotype). The availability of DNA markers polymorphic between inbred mouse strains now allows for direct selection for a genomic region in the development of a congenic strain. This marker-assisted approach has been successfully used to isolate the effects of individual loci in multigene mouse models of epilepsy, insulin-dependent diabetes, and systemic lupus erythematosus (FRANKEL *et al.* 1995; MOREL *et al.* 1996; YUI *et al.* 1996).

The construction of congenic strains is a lengthy undertaking that can require 3–5 years; repeated backcrosses and then intercrosses are necessary, and the generation time in mice is 8–9 weeks. However, construction of strains that are suitable for the fine-structure mapping of individual loci that contribute to a multigene trait can be accomplished in fewer generations (and potentially in less time) by using genome-wide typing to select mice for breeding. The term "speed congenics" has been used by LANDER and SCHORK (1994) to describe congenic strains developed in three to four backcross generations by marker-directed breeding. Here, we consider several strategies for selecting animals for breeding on the basis of their genotypes and determine the consequences of each strategy based on the number of chromosomes that

Corresponding author: Michael M. Weil, Department of Experimental Radiation Oncology, Box 66, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030.
E-mail: sa64001@odin.mdacc.tmc.edu

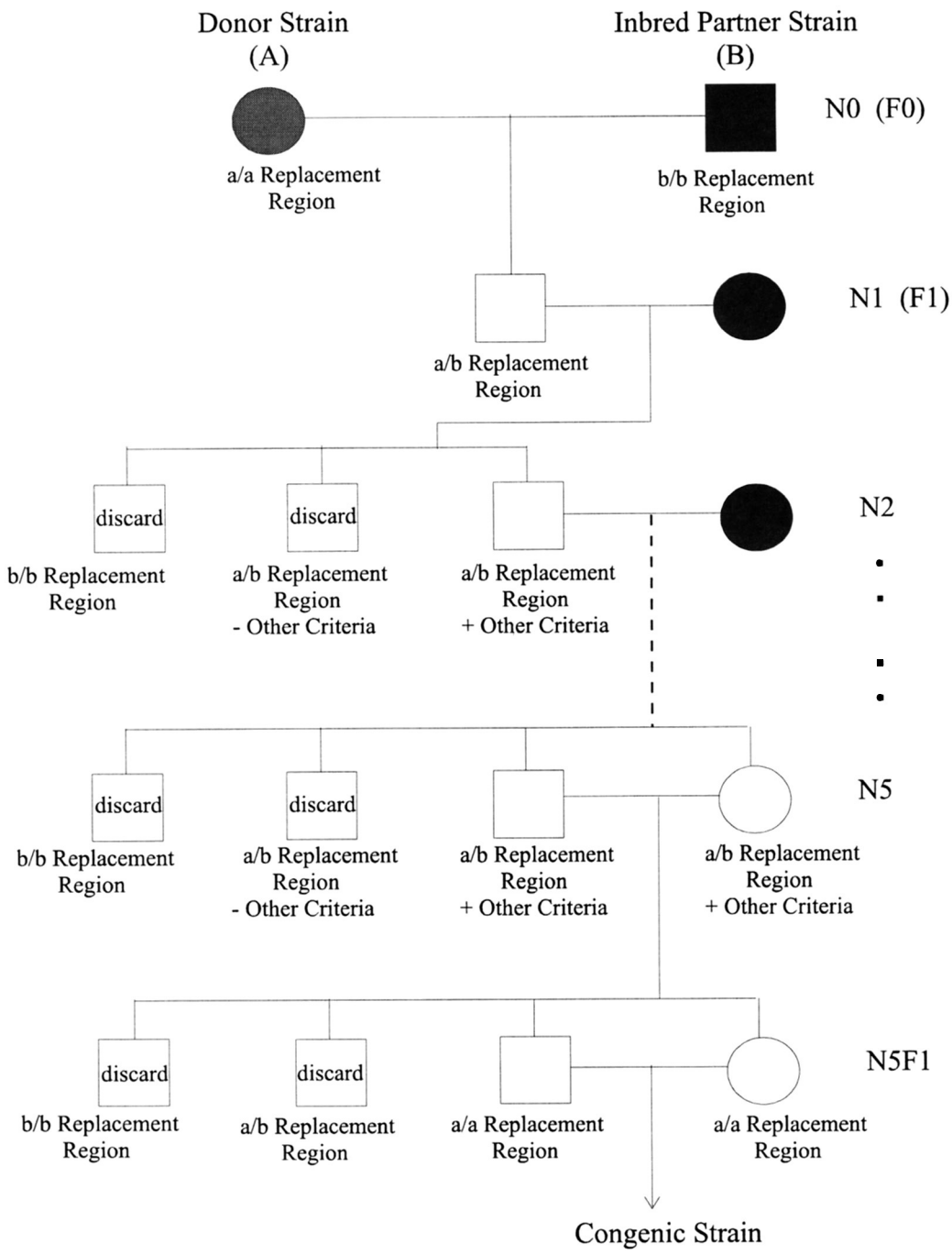

FIGURE 1.—Variation of a backcross breeding scheme to produce a congenic strain that has a strain A replacement region on a strain B genomic background. Progeny animals in each generation are backcrossed to strain B mice for several generations and then intercrossed to obtain offspring homozygous for the strain A replacement region. Polymorphic DNA markers spanning the replacement region are either entirely from strain A (a/a), entirely from strain B (b/b), or mixed (a/b). Other criteria are outlined by the strategies in MATERIALS AND METHODS.

contain donor genomic material outside of the replacement region and the average length of this unwanted donor DNA for each backcross generation. While this study pertains specifically to the mouse, many of the principles apply generally to other species.

MATERIALS AND METHODS

Summary of the rapid breeding method: The breeding scheme is a simple variant on the backcross procedure diagrammed in Figure 1. Initially, animals of the inbred partner strain are crossed with the donor strain and their F_1 progeny are backcrossed to the inbred partner strain. For the N_0 cross male inbred partner strain mice are used so that the inbred partner strain *Y* chromosome will be fixed in subsequent generations. Beginning at generation N_2, a male with the desired donor differential region is chosen to be repeatedly mated with females of the inbred strain until a specified number of N_3 male offspring is produced. One of these male offspring is chosen to father the next generation and the backcrossing continues as long as desired. A series of intercrosses is then used to generate animals homozygous for the donor-derived differential region.

The breeding partner from the backcross progeny is a male mouse for several reasons. In an attempt to control the positive and negative selection of the donor genome in each gen-

eration, only one mouse from the backcross progeny is used to found the next generation. The mouse chosen must produce a large number of offspring quickly, to allow selection of desired genomic characteristics for the parent of the succeeding generation. The choice of a male mouse as the breeding partner also insures that the congenic strain will have nonrecombined, inbred partner strain *X* chromosomes. Finally, the male autosomal genetic map tends to be shorter than the female autosomal genetic map and a lower recombination frequency in the noninbred partner improves the performance of the chromosome elimination strategies described below.

The factors that need to be determined are the numbers of male mice to produce in each generation and the strategy for selecting which male to use for breeding. The selection strategies considered for choosing the father of the next generation from those males having the desired differential region are presented below. All of the strategies are based on the selective elimination of the donor genome outside of the differential region. There are 19 autosomes in the mouse, and we assume in this evaluation that the differential region is located in one of them; calculations for the case in which there is more than one differential region and they are spread over several autosomes are not appreciably more difficult. Since the initial selection is for mice having the donor origin differential region, the donor origin chromosomes being eliminated in the strategies below are the 18 autosomes without the differential region. Sex chromosomes are not considered, since only a male is chosen to parent the next generation.

Methods for choosing the father of the next generation considered here are as follows:

Strategy 0: Make a totally random choice.

Strategy 1: Eliminate entire chromosomes containing donor origin sequence from the breeding population. Based on the genotyping results, choose the mouse with the fewest chromosomes containing donor origin material.

Strategy 2: If there is more than one suitable male according to strategy 1, calculate the sum of the lengths of the donor origin chromosomes that have been eliminated. Choose the male for which this sum is greatest to father the next generation.

Strategy 3: Sum the lengths of the donor origin chromosomes that have been eliminated and select the mouse for which this sum is greatest, but disregard the total number of donor origin chromosomes eliminated.

If, according to the breeding rule used, there are several males eligible to father the next generation, make a random choice between them.

Rules 2 and 3 can be justified on the grounds that as much undesired donor genome as possible should be eliminated in each generation. Also, long chromosomes are most liable to recombination, which may lead to interstitial regions of donor sequence on chromosomes in the next generation. Chromosome lengths in the mouse vary from 40.4 (chromosome *18*) to 115.5 cM (chromosome *1*), so summing the lengths of the inbred partner strain chromosomes gained in each of the backcross progeny weights the selection toward animals that have gained the most inbred partner strain material. It also weights the selection for mice who have gained longer inbred partner strain chromosomes, which are harder to transmit intact because of their propensity to undergo recombination.

Calculations: We use simulation to calculate, by generation of backcross, the probability of total elimination of specified numbers of donor strain origin chromosomes and of the mean length of the noneliminated donor segments in the remaining donor strain origin chromosomes.

Since our breeding strategies are based on selection at the chromosomal level, we will categorize chromosomes into one of two types. In the backcross breeding scheme, all progeny must have at least one chromosome in each chromosome pair derived entirely from the inbred partner strain. We designate this type as an IPT (inbred partner type) chromosome. The other chromosome in the homologous pair can either be derived entirely from donor strain sequences, entirely from inbred partner strain sequences or, because of a previous recombination event, contain both donor and inbred partner strain sequence. We term chromosomes that are of donor origin or have donor origin sequence as the result of recombination DMT (donor or mixed type) chromosomes. The breeding strategies considered select against transmission of DMT chromosomes.

Among the progeny of a backcross, a mouse will be homozygous for a particular inbred partner strain chromosome under three conditions. (1) Both parents are homozygous. Recall that the mother is always homozygous for any inbred partner strain chromosome and the father can either be homozygous for an inbred partner strain chromosome or heterozygous. Beginning at generation N_1, the breeding scheme makes it impossible for any animal to be homozygous for a donor strain chromosome. (2) During paternal gametogenesis there is an inbred partner strain chromatid that does not undergo recombination and the offspring receives this chromosome from the father. (3) A chromosome that contained both donor and inbred partner strain sequence undergoes a subsequent recombination that generates a chromosome derived entirely from inbred partner strain sequence, and this chromosome is inherited by the offspring.

We make simplifying assumptions detailed below to separate the simulation of recombination from that of the effect of breeding methods. This separation reduces the computer time required greatly. Initially we ignore the possibility that the genotyping analysis erroneously identifies a DMT chromosome as IPT. A later section discusses the effect of error.

Recombination: Two characteristics of successive recombinations of a DMT chromosome with its paired IPT type are used in this investigation: (1) the probability that a recombination will segregate the donor genome segments and so yield a pure IPT chromosome and a DMT chromosome, and (2) the length of donor genetic material remaining in the DMT chromosomes.

We use the term "state" to indicate the number of successive recombinations undergone by a DMT chromosome that has not become IPT through these recombinations. Chromosomes of offspring of the initial breeding of donor and inbred strains (generation N_1) have state 0. A recombination between an IPT chromosome and a DMT chromosome of state i yields either one IPT chromosome and a DMT chromosome of state $i + 1$ or else two DMT chromosomes of state $i + 1$. This observation does not imply that reciprocal products of recombination are likely to be recovered; they are not. Rather, this observation is used to calculate the probability distribution of chromosome types from which those in the offspring are drawn.

The simplifying assumption mentioned above is that all instances of a particular chromosome of the same state are identical in the probability that a recombination will produce a pure IPT chromosome and in the length of donor genetic material it contains. Thus, our simulation need only record the mean value of these characteristics by state; this results in a great savings of computer time.

To test the magnitude of the error introduced by this simplification, we compared results from the current method with those in which chromosome-by-chromosome simulation

was performed, and the simplification not used. For our standard case of using strategy 2 with 10 eligible male offspring per generation, these two calculations gave mean numbers of IPT chromosomes per generation that agreed to within 0.04, a negligible error.

We assume that recombination events are governed by the Haldane mapping function, which simplifies computations by erroneously assuming that there is no interference in recombination events. According to this function, the number of crossover points follows a Poisson distribution with mean L, the length of the chromosome in morgans. One morgan is that length of a chromosome within which, on average, one crossover occurs per gamete. According to this function, the probability of no crossover is $\exp(-L)$. Crossover points are distributed randomly (uniformly) over the length of the chromosome. If there are one or more crossovers, the genetic material of the resultant chromosomes is the same as their originators up to the first crossover point. From this point to the second crossover point or the end of the chromosome, the genetic material of the originators is transposed in the resultant chromosomes. The pattern continues, alternating identity and transposition through successive crossover points.

Simulation of successive recombination was repeated 10,000 times for each of the 19 autosomal mouse chromosomes. The probability by state of obtaining a pure inbred strain chromosome through the final recombination event is shown graphically in the top of Figure 2. The lines in Figure 2 show the results for the shortest chromosome (chromosome *18*, length 40.4 cM), the longest chromosome (chromosome *1*, length 115.5 cM), and a chromosome of mean length (72.1 cM). The figure shows that the probability that a recombination will yield one IPT chromosome increases with the number of recombination events; that is because the donor genomic material becomes successively diluted but generally remains concentrated in one contiguous segment of the chromosome. A recombination will yield two DMT chromosomes only if a crossover point falls within this short segment. The bottom of Figure 2 shows the proportion of donor chromosome material remaining at each state for DMT chromosomes. The mean proportion of donor genome decreases more slowly than a factor of $^1/_2$ because IPT chromosomes resulting from recombination do not enter into this average.

Simulation strategy: Recall the probability of no crossover is $p_n = \exp(-L)$, where L is the chromosome length in morgans. For each state, a, of a chromosome, we know from simulations the probability, p_a, that a crossover will result in one IPT chromosome and one DMT chromosome.

If one particular chromosome in the father is DMT, the probability that an offspring will receive an IPT chromosome from the father (it is sure to receive one from the mother) is

$$^1/_2(p_n + (1 - p_n)p_a)$$

Either there will be no crossover, an event of probability p_n, or there will be a crossover and the crossover will result in one chromosome of IPT, an event of probability $(1 - p_n)p_a$. The factor of one-half is the probability that the IPT chromosome will be the one contributed to the offspring.

The probability that an offspring will receive a chromosome in the same state as that of the father is

$$^1/_2p_n;$$

i.e., no crossover occurs, and the transmitted chromosome is the one that is not IPT.

The final possibility is that a crossover results in the transmission of a DMT chromosome of a state one larger than that of the DMT chromosome of the father. This probability is

$$(1 - p_n)(1 - p_a) + {}^1/_2(1 - p_n)p_a.$$

The first term is the probability that crossover results in two chromosomes in a state one larger than the DMT chromosome of the father; the offspring is certain to receive one of these. The second term is the probability that a crossover results in one IPT chromosome and one chromosome in a state one larger than the DMT chromosome of the father. The offspring receives the DMT chromosome with probability $^1/_2$.

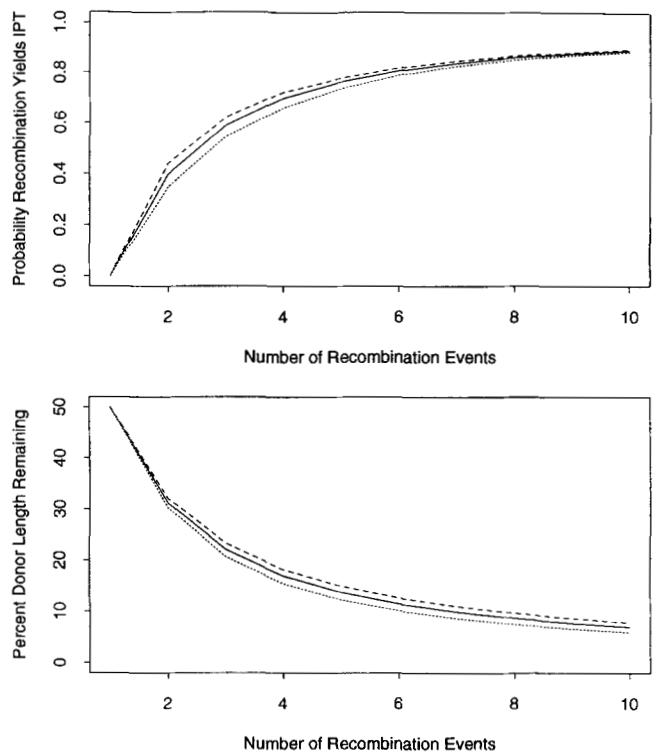

FIGURE 2.—(Top) Probability of a crossover between an inbred partner type (IPT) and donor or mixed type (DMT) chromosome yielding one chromosome that is IPT as a function of the total number of crossover events experienced by the DMT chromosome. (Bottom) Percentage of donor chromosome length remaining as the same function. —, a chromosome of mean length; – – –, the shortest chromosome; · · · · ·, the longest chromosome.

These formulas allow us to efficiently simulate offspring in a backcross with a father whose chromosomes are in any particular configuration of states. The simulation must consider each chromosome separately, since the chromosome length affects the values of the p's. Breeding strategy simulations were also replicated 10,000 times to achieve an overall estimate of the effect of each breeding method.

Software availability: A standard conforming Fortran 90 program for performing the calculations described here is available by anonymous ftp to odin.mdacc.tmc.edu (see file "readme" in ./pub/congenic). The program is available both in source form and as a DOS self-extracting executable.

The program calculates by simulation the distribution of IPT chromosomes remaining and the mean length of donor chromosome material by generation for a specified number of eligible male offspring and a specified breeding strategy. This program allows the differential genomic region of inter-

TABLE 1

Cumulative distribution of DMT chromosomes remaining

Cumulative number	N_2	N_3	N_4	N_5	N_6	N_7
13	0.9984	1.0000	1.0000	1.0000	1.0000	1
12	0.9617	1.0000	1.0000	1.0000	1.0000	1
11	0.7609	1.0000	1.0000	1.0000	1.0000	1
10	0.4224	1.0000	1.0000	1.0000	1.0000	1
9	0.1631	1.0000	1.0000	1.0000	1.0000	1
8	0.0467	0.9999	1.0000	1.0000	1.0000	1
7	0.0121	0.9875	1.0000	1.0000	1.0000	1
6	0.0027	0.9077	1.0000	1.0000	1.0000	1
5	0.0002	0.6644	1.0000	1.0000	1.0000	1
4	0.0000	0.3437	0.9997	1.0000	1.0000	1
3	0.0000	0.1143	0.9853	1.0000	1.0000	1
2	0.0000	0.0224	0.8544	1.0000	1.0000	1
1	0.0000	0.0029	0.4756	0.9933	1.0000	1
0	0.0000	0.0002	0.1039	0.8283	0.9973	1
			Mean number of retained DMT chromosomes			
	10.6319	4.9570	1.5811	0.1784	0.0027	0
			Percent donor genetic material			
	22.07	8.19	2.17	0.21	0.00	0.00

est to be spread over several chromosomes; it need not be concentrated on one chromosome as in the example in this paper.

RESULTS

The simulations are based on the development of a congenic pair having a differential region on chromosome *16*, so donor material on chromosomes *1–15* and *17–19* is targeted for elimination. Strategy 2 with 10 eligible male offspring is chosen as the baseline against which changes are measured. Strategy 2 is the best strategy of the three, although by an inappreciable amount.

Operating characteristics of the baseline strategy: The operating characteristics are shown in Table 1. Chromosome *16* is excluded from consideration since it contains the differential region and consequently is not targeted for elimination. The top of the table gives the probability of a particular number or fewer DMT chromosomes by generation. For example, the probability of six or fewer DMT chromosomes in the N_3 generation is 0.9077. The middle of the table shows the mean number of DMT chromosomes by generation, and the bottom shows the mean proportion of the genome composed of donor type sequence.

A column providing the characteristics of the N_1 generation is not shown because it is known with certainty that all autosomal pairs consist of one IPT and one DMT chromosome. The rows of Table 1 corresponding to the probability of 18 . . . 14 or fewer retained IPT chromosomes are not shown because this probability is uniformly 1 for generations of N_2 or later.

Graphs of the mean number of retained DMT chromosomes and the mean proportion of the total genetic material from the donor strain are shown in Figure 3 as functions of the breeding generation.

Effect of breeding rule: Figure 4 shows the effect of the breeding strategy as differences from the baseline operating characteristics, *i.e.*, the characteristics of strategy 2 are forced to be zero. From this figure, it is evident that strategies 1, 2, and 3 produce much the same operating characteristics. The largest difference in mean number of DMT chromosomes eliminated between the three strategies is 0.07.

It is also evident that the selection by genotype of the male mouse used in the backcrosses is a far more efficient method for rapidly generating congenic pairs than the random choice of father (strategy 0). In the N_4 generation, the other strategies eliminated an average of more than 4.5 DMT chromosomes more than does strategy 0. The proportion of the total genetic material from the donor strain shows very similar behavior for strategies 1–3, all are much better than strategy 0.

As far as we know, this work provides the first examination of the number of IPT chromosomes by generation under random inbreeding; previous authors used only a successive dilution argument on donor material length.

Effect of number of eligible offspring: Figure 5

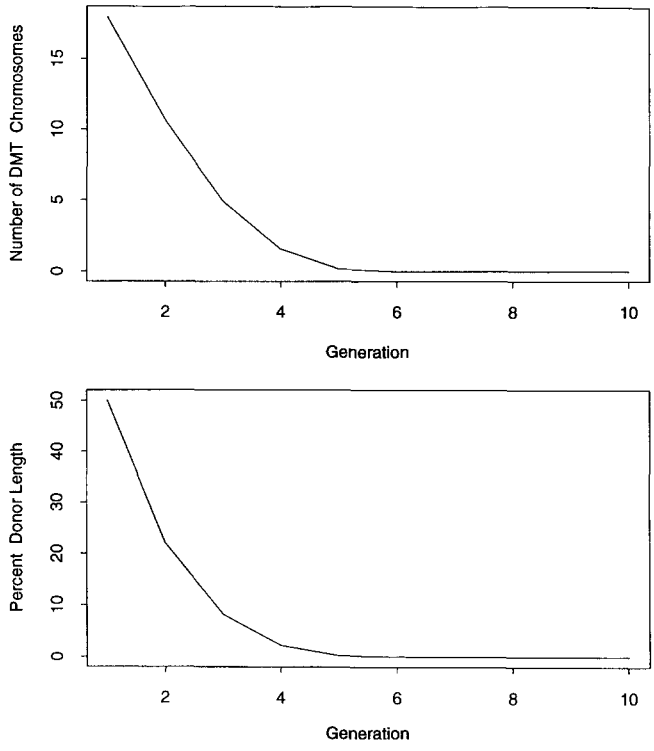

FIGURE 3.—(Top) Mean number of donor or mixed type (DMT) chromosomes remaining by generation of backcross using strategy 2 with 10 eligible offspring per generation. (Bottom) Percentage of donor genetic material remaining by generation for the same breeding strategy. Generation 1 on this and all following figures is generation N_1 of Figure 1, generation 2 is N_2, and so on.

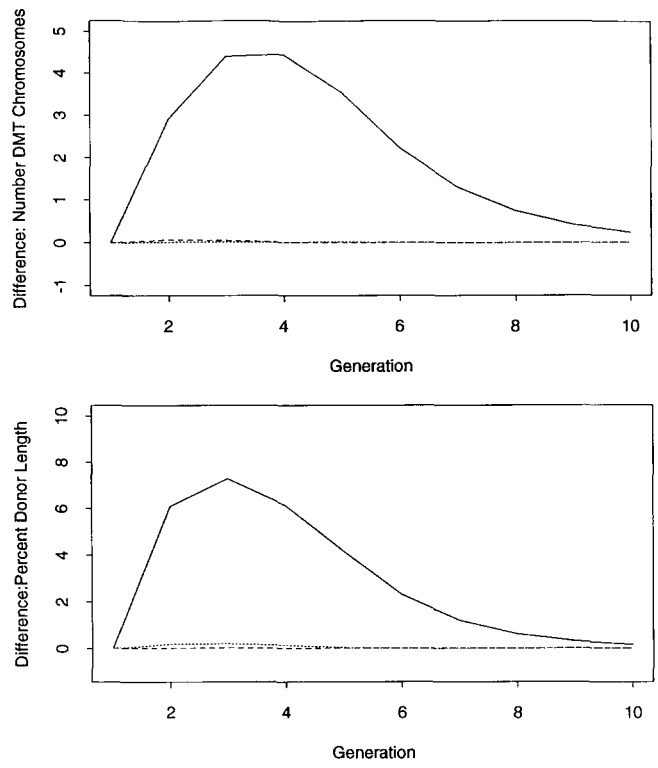

FIGURE 4.—Differences from results obtained under the standard condition: breeding rule 2 with 10 eligible offspring per generation. (Top) Average number of DMT chromosomes for strategies 0, 1, and 3 with 10 eligible offspring per generation minus the average for the strategy 2. (Bottom) Average length of donor genetic material retained for the same breeding rules minus that for the standard condition. In both panels, smaller numbers represent better performance. —, difference for strategy 0; · · · · ·, strategy 1; – – –, strategy 3. Strategies 1–3 are seen to produce almost identical results, which are much better than those of strategy 0 (random breeding).

shows the difference in outcome, using strategy 2, from breeding until five, 15, and 20 eligible male offspring (*i.e.*, those carrying a copy of the donor origin differential region) are available instead of the baseline number of 10. As expected, more offspring add to the effectiveness of the breeding scheme although there is a diminishing return. In the N_3 generation, 10 offspring result in 1.1 fewer DMT chromosomes than do five; the change from 10 to 15 offspring is 0.54; from 15 to 20, it is 0.35 chromosomes.

Perhaps a more useful gauge of the effect of the number of offspring is the probability that one or zero DMT chromosomes remain in generation N_4. For five offspring, these probabilities are 0.15 and 0.03, respectively; for 10 offspring, they are 0.37 and 0.11; for 15, 0.49 and 0.21; for 20, 0.52 and 0.32.

Error in screening: We next examined the effect of error in screening chromosomes. All simulations described above were repeated with 5-cM screening intervals, and the consequent possibility that a DMT chromosome could be incorrectly identified as being IPT was considered.

The number of markers for a particular chromosome is obtained by dividing its length in cM by five and discarding any fractional part of the answer. The screening markers are assumed to be present at both endpoints of the chromosome and to be equally spaced between these endpoints.

The modifications to the simulation to accommodate error were minor. We say that a DMT chromosome is missed if the screening fails to detect the donor genetic material; hence the chromosome is falsely deemed to be IPT. The property of whether each chromosome would or would not be missed in screening was recorded in addition to state. Recombination events between a chromosome that would be missed in screening with an IPT must yield either two chromosomes that would be missed or one IPT chromosome and one that would be missed. To keep the computation to a manageable level, the assumption was made that the probability of a recombination between a missed chromosome in a particular state yielding one IPT chromosome and one that would be missed is the same regardless of the state at which the original chromosome became one that would be missed.

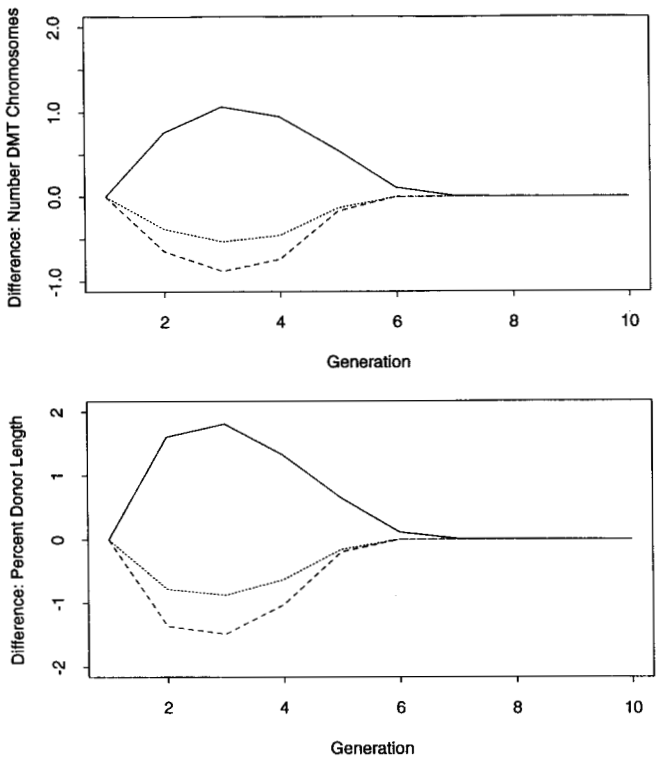

FIGURE 5.—Effect of number of eligible offspring. Shown are differences from results obtained under the standard condition: breeding rule 2 with 10 eligible offspring per generation. (Top) Difference in average number of donor or mixed type (DMT) chromosomes retained by generation. (Bottom) Difference in percentage of the length of donor genetic material retained. —, five eligible offspring; · · · · ·, 15; – – –, 20.

Properties of missed DMT chromosomes: Assuming that the genotypic analysis uses markers placed at 5-cM intervals, the mean length of the total donor genome in the chromosomes that would be missed in screening is very low, with a mean of 1.5–2 cM regardless of state and chromosome number. Because this length is so short, a recombination between a missed chromosome and an IPT chromosome almost always results in one IPT chromosome; the probability is greater than 0.95 in all cases considered. To produce two DMT chromosomes, a crossover point would have to fall within the short donor genome region.

The probability that recombination will produce a chromosome that will be missed increases steadily with the state of the chromosome. For state 1, the probability ranges between 0.006 and 0.01 depending on the length of the chromosome; for state 9, the range is 0.25–0.43. There are two reasons for this increasing proportion. One is that the length of donor genome material in a chromosome decreases with state, thus producing shorter segments. The other reason is that the chromosomes that would be missed accumulate over successive recombination events.

Effect of error in screening: Fortunately, the effect of screening error is quite small when markers at 5-cM intervals are used. The maximum (over all generations of all simulation cases) average number of missed chromosomes is only 0.21 and the maximum average missed donor genome length is 0.34 cM. Of course, the latter figure averages the good chance that no DMT chromosome is missed in screening with the smaller probability that a donor genome stretch of length 1.5–2.0 cM is missed.

Effect of avoidance region: Crossover points in recombinations tend not to be too close together, a factor not considered in the previous simulations. To determine the effect of this consideration, we repeated the previous simulations with an avoidance region of 10 cM.

In simulating recombination with an avoidance region, the number of crossover points is generated from a Poisson distribution as before. As each successive crossover point is placed randomly along the length of the chromosome, an avoidance region of 10 cM is drawn in either direction from the crossover position. Should this region include another crossover point, a different random placement is attempted. This process is continued until there are 10 failures to place the point, then the last point generated is used.

The chromosomes generated using this version of avoidance regions almost always had recombination points at least 10 cM apart. The worst violator of this interference was on the shortest chromosome (40.4 cM). Violations occurred in 537 of 100,000 simulated cases in which at least one recombination occurred; this gives a rate of 178 per 100,000 for simulated meioses whether or not there was a recombination. For the average size chromosome (72.1 cM), the rates were 122 and 63 per 100,000, respectively. For the longest chromosome (115.5 cM), the rates were 25 and 17 per 100,000.

The effect of the avoidance region is to lessen the proportion of short stretches of donor genome material. However, the results changed very little from those of the previous section. The maximum average number of missed chromosomes decreased from 0.21 to 0.15 and the maximum average missed donor genome length decreased from 0.34 to 0.25.

DISCUSSION

The simulations we have run show that by selecting animals for breeding on the basis of their genotypes it is possible to significantly compress the numbers of generations and potentially the time scale required for making congenic pairs suitable for high-resolution mapping. We compared several strategies based on eliminating entire chromosomes of donor origin from the breeding population. Of these strategies, the most effective is to backcross the mouse from each progeny generation that has the fewest chromosomes containing donor origin material. If there is more than one suitable

mouse, sum the lengths of the donor origin chromosomes that have been eliminated and select the mouse for which this sum is greatest.

Two routes can lead to the elimination of donor origin sequences from the breeding population. Progeny mice always receive a haploid genome equivalent from the inbred partner used as the mother in each backcross. The germline of the mouse used to father the backcross always has a haploid genome equivalent derived from the inbred partner strain; the other half of the genome is either entirely from the donor strain (generation N_1) or a mixture of donor and inbred partner sequences (generation N_2 and on). Donor sequence elimination occurs if a pup receives a paternal IPT chromosome that has never contained donor sequences (first route) or the pup receives a paternal chromosome that has both donor and inbred partner sequence as a result of recombination (second route). In the second route, the donor sequence may be partly or entirely lost by recombination in subsequent generations. In practical terms, the second route speeds up the process of donor sequence elimination as shown in the top of Figure 2, but complicates the genotyping since it can generate progressively smaller fragments of donor sequence that will eventually escape detection as shown in the bottom of Figure 2.

In actual practice other factors will come into play when deciding on a strategy. The object of constructing congenic pairs is to eliminate the influence of the background genome on the trait under investigation. In some cases other loci that contribute to the trait will have already been identified during initial mapping studies. Elimination of the donor strain alleles of these other loci is particularly important. As the generation number increases, the size distribution of the donor strain fragments generated by recombination decreases and, consequently, the risk of missing donor-derived sequence in the genotyping is more likely in the later generations. A modified strategy might weight the selection criteria to eliminate chromosomes containing unwanted loci in the earliest backcrosses.

The efficiency of the chromosome elimination strategy improves as the number of progeny to select a breeder from increases. Figure 5 illustrates that increasing the number of eligible mice from five to 10 greatly improves elimination of donor origin chromosomes, but that additional increases have progressively less effect. Large numbers of potential breeders are most advantageous at generation N_3. It should be kept in mind that the mice referred to in Figure 5 are male and contain the differential region from the donor strain. What does this mean in regard to the numbers of progeny that must be available in each generation? The probability of a mouse being male is 0.5. Assuming a 26-cM differential region, the probability that a mouse would receive an intact donor sequence for that region from a heterozygous parent is $0.5 \exp(-0.26)$, or ~ 0.38. To have 10 mice eligible for breeding requires a population of ~ 50 animals.

Generating large numbers of progeny from a young, partially inbred male mouse and then genotyping these progeny adds to the generation time. When breeding progeny populations of 50–100 animals, three factors are important: numbers of litters, litter size and surviving pups. The number of litters can be maximized by rotating the male through cages containing two females at 24-hr intervals. Litter sizes vary by strain, but strain choice might not be under investigator control. In some strains pup survival is determined by the breeding history of the female used in the mating. For example, C57BL/6 females often partially or completely cannibalize their first couple of litters, but raise succeeding litters without loss of pups. In these strains, females that have previously raised large litters are a better choice for breeding than inexperienced animals. In cases where litter losses are high and not enough experienced females are available, fostering the pups onto more suitable females of a different strain is possible.

No mention has been made of selection against donor strain material on the same chromosome as the replacement region but outside of the region that contains it. The probability that donor sequences remain decreases with their distance from the selected region. If another locus that influences the trait under study has been mapped to this chromosome, the selection criteria can be modified to eliminate this passenger locus. Also, if the size of the unselected donor material adjacent to the selected differential genomic region can be fortuitously minimized during the construction of the strains, subsequent mapping may be simplified.

We are currently trying to determine the most efficient approach for genotyping the mice. One approach that would minimize the numbers of PCR reactions required would be to screen markers in succeeding tiers. For example, in screening generation N_2 mice a first tier screen might be for markers flanking the replacement region. Mice not having a donor-derived replacement region need not be screened further. The second tier screen would be for markers at 5-cM intervals in the replacement region to assure that a double recombination event had not occurred in that region. For the third tier, the most proximal and distal marker on each of the other 18 autosomes would be tested. Only a few of the mice identified in the third tier as having lost the most DMT chromosomes need be tested at the 5-cM level, and then only on those chromosomes in which both flanking markers are derived only from the inbred partner strain. With each succeeding generation, fewer assays are required because once donor-derived sequences are no longer found on a chromosome there is no need to screen that chromosome in subsequent generations.

We thank Dr. BENJAMIN A. TAYLOR for helpful suggestions and for a critical review of this manuscript and Mr. WALTER J. PAGEL for editorial assistance. This work was supported in part by grants CA-16672 and CA-06294 from the National Cancer Institute, a cooperative study agreement with IBM, and by the personal generosity of LARRY and PAT MCNEIL.

LITERATURE CITED

DIETRICH, W., H. KATZ, S. E. LINCOLN, H. S. SHIN, J. FRIEDMAN *et al.*, 1992 A genetic map of the mouse suitable for typing intraspecific crosses. Genetics **131:** 423–447.

FRANKEL, W. N., E. W. JOHNSON and C. M. LUTZ, 1995 Congenic strains reveal effects of the epilepsy quantitative trait locus, El2, separate from other El loci. Mamm. Genome **6:** 839–843.

LANDER, E. S., and N. J. SCHORK, 1994 Genetic dissection of complex traits. Science **265:** 2037–2048.

LINCOLN, S., M. DALY and E. LANDER, 1992 Mapping genes controlling quantitative traits with MAPMAKER/QTL 1.1. Whitehead Institute Technical Report, Ed. 2, Cambridge, MA.

MOREL, L., Y. YU, K. R. BLENMAN, R. A. CALDWELL and E. K. WAKELAND, 1996 Production of congenic mouse strains carrying genomic intervals containing SLE-susceptibility genes derived from the SLE-prone NZM2410 strain. Mamm. Genome **7:** 335–339.

PATERSON, A., E. LANDER, S. LINCOLN, J. HEWITT, S. PETERSON *et al.*, 1988 Resolution of quantitative traits into Mendelian factors using a complete RFLP linkage map. Nature **335:** 721–726.

YUI, M. A., K. MURALIDHARAN, B. MORENOALTAMIRANO, G. PERRIN, K. CHESTNUT *et al.*, 1996 Production of congenic mouse strains carrying NOD-derived diabetogenic genetic intervals—an approach for the genetic dissection of complex traits. Mamm. Genome **7:** 331–334.

Communicating editor: C. KOZAK