Theoretical and empirical issues for marker-assisted breeding of congenic mouse strains

Paul Markel[1], Pei Shu[1], Chris Ebeling[2], George A. Carlson[2], Deborah L. Nagle[1], John S. Smutko[1] & Karen J. Moore[1]

Congenic breeding strategies are becoming increasingly important as a greater number of complex trait linkages are identified. Traditionally, the development of a congenic strain has been a time-consuming endeavour, requiring ten generations of backcrosses. The recent advent of a dense molecular genetic map of the mouse permits methods that can reduce the time needed for congenic-strain production by 18–24 months. We present a theoretical evaluation of marker-assisted congenic production and provide the empirical data that support it. We present this 'speed congenic' method in a user-friendly manner to encourage other investigators to pursue this or similar methods of congenic production.

Mouse models continually support the successful genetic analyses of complex human diseases. Many of these models centre on unravelling the genetic differences between the inbred strains of mice. For example, several linkages of quantitative trait loci (QTLs) have already been detected in mice for such complex conditions as alcohol sensitivity, atopy, diabetes, autoimmunity and obesity[1–7]. This current wave of dissecting complex diseases in the mouse has brought about the need for efficient ways of moving (introgressing) genomic sections from one strain to another, creating congenic strains[6,7]. Congenic strains are traditionally developed by continually backcrossing a donor strain that harbours a gene or genomic region of interest with a recipient inbred strain[8,9]. According to Mendel's laws, it is expected that, on average, half of the unrelated genomic material will be transmitted to a subsequent backcross generation. Ten generations of backcross matings result in mice heterozygous for the locus of interest that are approximately 99.90% recipient genome and carry an introgressed region of interest[8,9]. The introgressed region is then made homozygous by intercrossing, resulting in a congenic strain. This congenic strain theoretically represents a genetic background identical to that of the recipient inbred strain with the exception, on average, of the 20-cM introgressed donor genomic region of interest. This traditional method of congenic production can take upwards of 2.5–3 years to complete.

The speed and efficiency of developing congenic strains are now greatly enhanced with the recent completion of a microsatellite-based murine map. Microsatellite markers can be used to follow the inheritance of the region(s) or gene of interest. In this fashion, the researcher is not entirely dependent upon phenotypic information in the selection of appropriate breeders. Moreover, the optimal breeder can be selected on the additional criterion that it harbours the least amount of donor genome. The latter technical advantage effectively reduces the number of generations necessary to construct a congenic strain. We have used such marker-assisted congenic-strain production to produce a series of congenic strains harbouring the *Apoe* null allele[10]. The *Apoe* knockout mouse offers a promising model for study of the genetic components in the development of atherosclerosis[10–12]. Animals homozygous for the *Apoe* knockout spontaneously develop a wide range of lesions on a normal diet. We have undertaken the production of several congenic lines of mice carrying the *Apoe* knockout in order to identify QTLs that modulate the development of atherosclerotic lesions.

Here we discuss the theoretical and empirical issues concerning marker-assisted development of congenic strains. In the first section, we develop the statistical expectations for a marker-assisted congenic breeding design. In the following section, we compare the theoretical expectations with the empirical results of a marker-assisted breeding

Table 1 • Traditional congenic breeding strategy

Generation	Average % heterozygous (D/R) segments ± SD	% recipient genome
F_1	100.00	50.00
N_2	50.00 ± 7.07	75.00
N_3	25.00 ± 5.00	87.50
N_4	12.5 ± 3.54	93.75
N_5	6.25 ± 2.5	96.88
N_6	3.13 ± 1.76	98.44
N_7	1.56 ± 1.25	99.22
N_8	0.78 ± .88	99.61
N_9	0.39 ± 0.63	99.81
N_{10}	0.20 ± 0.44	99.90

Table 2 • Marker-assisted congenic breeding strategy

Backcross generation	Average % D/R segments ± SD	% D/R segments in 'best' male	% recipient genome of 'best' male
F_1	100± 0	100	50
N_2	50.00 ±7.07	38.32	80.84
N_3	19.16 ±4.38	11.93	94.03
N_4	5.98 ±2.44	1.95	99.03
N_5	0.98 ±0.98	~0	~100

The prediction is based on 20 male carriers of the target gene screened. Hence, 1.65 standard deviations from the mean can be acheived. Each generation is propagated from the 'best' male of the previous generation.

[1]*Millennium Pharmaceuticals, 640 Memorial Drive, Cambridge, Massachusetts 02139, USA.* [2]*The McLaughlin Research Institute, Great Falls, Montana 59405, USA. Correspondence should be addressed to K.J.M. e-mail: moore@mpi.com*

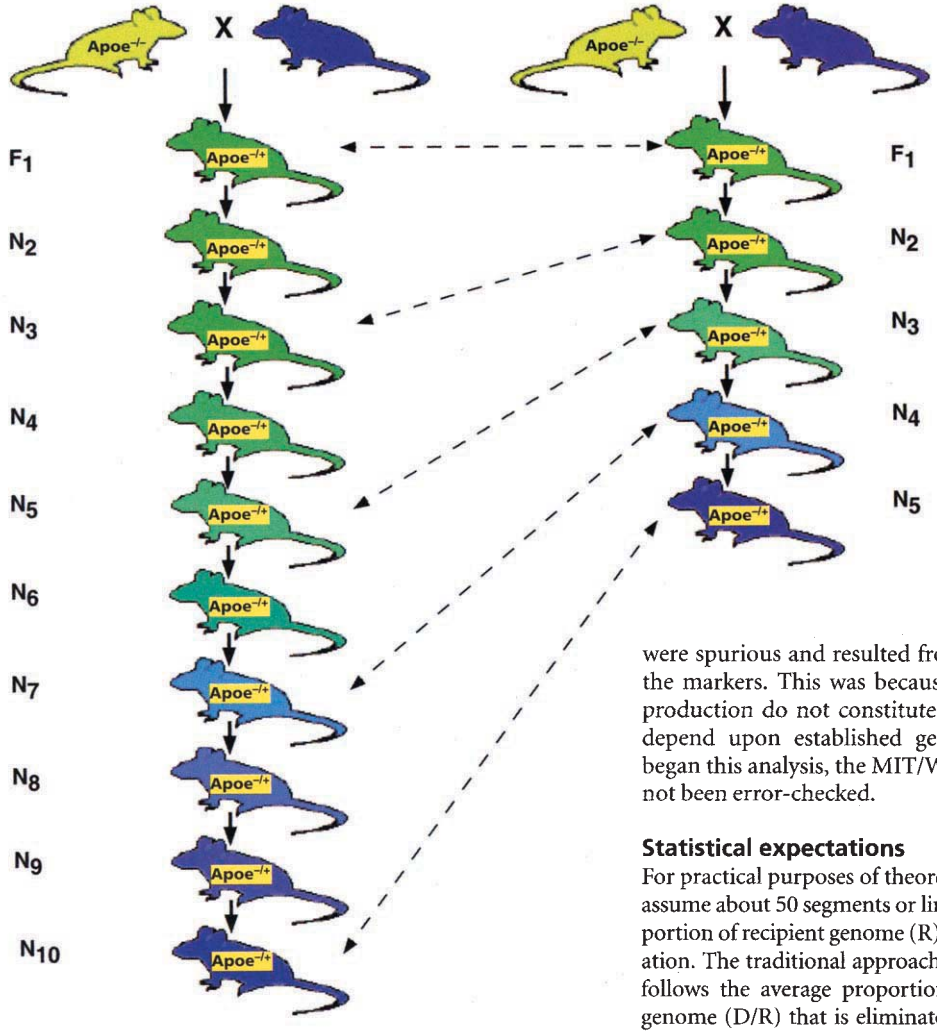

Fig. 1 This cartoon shows the production of a congenic strain using the conventional ten generations of backcross (left) *versus* the 'speed' congenic method described in this paper (right). Yellow indicates the donor genome, blue the recipient genome. The yellow boxes containing '*Apoe*' represent the target region of introgression.

of the *Apoe* null allele congenic strains. Our approach is straightforward, lending itself to practical application and interpretation for a real laboratory setting.

Marker development

We carried out marker evaluation to compare allele size between the 129/SvJ strain and each of the following inbred strains: C57BL/KsJ, BALB/cByJ, C3H/HeJ, DBA/2J, FVB/NJ and CAST/Ei. Our goal was to obtain a series of polymorphic markers at approximately 10-cM intervals that were not subject to preferential amplification. A set of such markers was established for each strain pair. In light of recent papers that describe the differences between the various 129 inbred mouse strains[13,14], 129/SvJ DNA was not ideal to for marker development, as we intended to work with a 129/Sv ter-*Apoe* donor strain. However, we did not find that many of our chosen markers were negated during the production of the congenics because of the 129 substrain differences.

Using these markers, we performed genome scans on all N_2 generation mice at approximately 20-cM intervals. We also scanned the subsequent N_3 generation at 20-cM intervals, but the set of markers was interspersed between the markers used in the N_2 generation. In effect, this results in an approximately 10-cM interval genome-wide scan. Additionally, it facilitated a finer assessment of double recombinants. A few true double recombinants were detected at the N_3 analysis stage, but most of those were spurious and resulted from incorrect genetic map order of the markers. This was because genome scanning and congenic production do not constitute a genetic mapping technique but depend upon established genetic maps. Moreover, when we began this analysis, the MIT/Whitehead murine genetic map had not been error-checked.

Statistical expectations

For practical purposes of theoretical computations (Tables 1,2), we assume about 50 segments or linkage groups for measuring the proportion of recipient genome (R) retained with each backcross generation. The traditional approach for congenic strain development[8,9] follows the average proportion of heterozygous donor/recipient genome (D/R) that is eliminated with each round of mating. The percentage of D/R is reduced, on average, by 50% with each subsequent generation (Table 1).

The expected change in genomic material for a backcross mating can be predicted with a normal approximation of the binomial distribution. The binomial distribution in this case indicates that each backcross mating can result in linkage groups that either remain heterozygous (D/R) or are fixed homozygous for the recipient strain (R/R). Let n=D/R segments for generation 'N', p=probability that a D/R segment is fixed R/R and q=probability that a D/R segment remains heterozygous. The average number of D/R segments in the subsequent generation is equal to $\mu_{(N+1)}=\mu_N p$. The standard deviation is equal to $\sigma_{(N+1)}=(\mu_N pq)^{1/2}$.

Note that the standard deviation decreases exponentially with each backcross generation (Table 1). From the number of D/R segments one can calculate the percent recipient genome: (100% – 1/2 (%D/R)).

Table 3 • Theoretical potential

Number of male carriers	Potential reduction in D/R (standard deviation)
5	0.85
10	1.29
15	1.50
20	1.65
30	1.84
40	1.96
50	2.06

Table 4 • Empirical data

Recipient strain	Backcross generation	Estimated avg. % D/Ra	No. Male carriers	Potential gain (z)[b]	Estimated % D/R for best male[c]	Estimated % recipient genome for best male[d]	Observed avg. % D/R among male carriers	Observed % D/R for best male[e]	Observed% recipient genome for best male
BALB/cByJ	N_2	50.00	4	0.68			39.96±9.24	27.27	86.36
	N_3	13.64±3.21	10	1.29	9.50	95.25	10.60±1.26	7.69	96.15
								(10.26)	(94.87)
	N_4	5.13±1.48	14	1.48	2.94	98.53	5.71±1.17	3.82	98.09
	N_5	1.91±0.78	9	1.22	0.96	99.52	2.98±0.56	1.79	99.11
C3H/HeJ	N_2	50.00	17	1.56			44.21±7.73	31.88	84.06
	N_3	15.94±3.40	8	1.13	12.10	93.95	13.73±3.64	8.48	95.76
	N_4	4.24±1.38	10	1.29	2.46	98.77	8.52±2.21	4.32	97.84
	N_5	2.16±0.82	5	0.85	1.46	99.27	3.00±0.94	1.18	99.41
C57BL/Ks	N_2	50.00	8	1.13			43.41±4.34	37.92	81.04
	N_3	18.96±2.52	7	1.07	16.26	91.37	18.31±3.77	12.66	93.67
	N_4	6.33±1.42	5	0.85	5.12	97.44	6.57±2.53	2.07	98.96
	N_5	1.04±0.55	4	0.68	0.67	99.66	1.55±0.64	0.59	99.70
CAST/Ei	N2	50.00	5	0.85			45.45±6.43	36.36	81.82
								(42.42)	(78.79)
	N3	21.21±4.01	21	1.67	14.51	92.74	17.35±3.64	8.93	95.54
	N4	4.47±1.41	1	—	—	—	8.24	8.24	95.88
	N5	4.12±1.10							
DBA/2J	N_2	50.00	17	1.56			45.59±6.80	34.17	82.92
								(40.00)	(80.00)
	N_3	20.00±4.08	2	0	20.00	90.00	20.63±0.63	20.00	90.00
	N_4	10.00±1.77	14	1.48	7.38	96.31	10.20±2.60	6.56	96.72
	N_5	3.28±1.01	9	1.22	2.05	98.97	3.49±1.32	1.24	99.38
FVB/NJ	N_2	50.00	13	1.41			45.05±5.59	35.71	82.14
								(38.10)	(80.95)
	N_3	19.05±3.89	16	1.56	12.98	93.51	12.35±3.16	7.77	96.12
	N_4	3.89±1.37	12	1.39	1.99	99.00	5.21±1.41	3.43	98.29
	N_5	1.72±0.70	4	0.68	1.24	99.38	1.22±0.59	0.54	99.73

[a]Estimated average % D/R is estimated on the basis of the results of the 'best' male from the previous generation. The standard deviation is based on the binomial distribution for the particular number of markers used in the previous generation, and is converted to a percentage of those markers. [b]This value is a function of the number of available male carriers and is expressed as 'z', assuming a standard unit distribution. This value of z reflects the potential reduction in donor genome by sampling a single extreme male. [c](estimated average % D/R) – [(standard deviation)(z)]. [d]100% – (estimated % D/R for best male). [e]Parentheses indicate the percentage of recipient genome of the male used to breed the subseqent generation when the 'best' male was not used.

Conventional congenic-strain production requires ten generations of backcrossing and should achieve a strain carrying 99.90% recipient genome (Table 1). It can take upwards of 2.5–3 years of breeding to accomplish this. However, the number of generations necessary for developing a congenic strain can be reduced if a mouse is selected at each generation that retains significantly fewer D/R segments than expected by chance and is then used to propagate the next generation. With polymorphic DNA markers across the whole genome, the heterozygosity for (D/R segments) of each backcross mouse can be monitored. The male carrier with the lowest number of D/R segments is selected. For practical breeding purposes, we mate a male so that efficient use is made of the chosen 'best genotype'—that is, the male is mated to multiple recipient females. Clearly, this male also needs to be carrying the target gene from the donor strain. In general, each mouse represents 1/n of the normal distribution. The expected number of male carriers in each generation is equal to 0.25 n (that is, half the offspring are male and half of those carry the target gene). Hence, provided that the male can produce 80 offspring, there will be twenty male carriers of the subsequent generation to be screened for D/R distribution. Assuming a normal distribution of D/R segments among the progeny, it follows that the ideal male (retaining the fewest D/R segments) comprises 1/0.25 n of the lower tail of a normal distribution. With a standard normal distribution, this percentile is represented by a standard deviation (expressed in standard units 'z') from the average retention of D/R segments for a particular generation. Again, starting with 80 progeny, this percentile is 5%, which is equivalent to a standard deviation of 1.65. Thus, starting with 80 offspring, $(0.25)\times(0.05)$n mice should result in one ideal male breeder retaining 1.65 standard deviations fewer D/R segments than expected on average for each generation. With this scheme of breeding and DNA marker–assisted selection, the number of generations theoretically necessary to produce a congenic strain is five. The theoretical figures for the scheme described above are shown in Table 2.

For practical reasons, it is not always experimentally possible to produce twenty male carriers, the target number that supports an opportunity for selecting the ideal male carrier that has 1.65 standard deviations less D/R than the mean. Instead, each backcross mating will produce a particular number of male carriers from which the 'best' male carrier will be selected (that is, the one that carries the fewest D/R segments). For these practical reasons, the expectated maximal reduction of D/R segments needs to be estimated a new in each backcross generation on the basis of the actual number of males obtained. Potential results for donor-genome reduction are presented in Table 3 as a function of incrementally different numbers of progeny.

Theoretical expectations and empirical results compared

We have put the above theory into practice in the production of

multiple congenic strains carrying the *Apoe* gene knockout allele[10]. During the production of these congenic strains, each backcross mouse was genome-scanned at an effective marker density of approximately 10-cM intervals. The results obtained can be seen in Table 4. The first column of this table indicates the recipient inbred strain to which the *Apoe* knockout allele was transferred. The second column indicates the generation of backcrossing. The third column is the estimated average percent of D/R genome, based on the male from the previous generation that was used for propagation. The standard deviation associated with this mean is based on the number of markers yet to be fixed (that is, to be made R/R) in that male carrier. The fourth column represents the number of male carriers genome scanned for each generation. The fifth column indicates the potential reduction in D/R genome (expressed in standard units, 'z'), based on the number of male carriers for a particular generation. The next two columns represent the estimated percent D/R, and consequential percentage of recipient genome, for the 'best' male, as a function of columns 3 and 5 (see below). The final three columns represent the actual observation of heterozygosity for male carriers genotyped for a particular generation.

Depending on the number of male carriers genome-scanned in each generation, we multiply the standard deviation (given in the third column) by the potential gain (z, fifth column), estimating the percent of recipient genome of the potentially 'best' male breeder. We compare our estimate of the potentially 'best' male to the observed data (that is, the estimated percentage of recipient genome for 'best' male *versus* actual percentage of recipient genome for 'best' male). For example, in the recipient strain BALB/cByJ at generation N2, the actual 'best' male is 86.36% recipient genome. As a result, we expect the N_3 estimated average percentage of recipient genome to be $100-(100-86.36)/2=93.2\%$ (which is equivalent to 13.64% D/R) and that of the 'best' male to be 95.25% recipient genome if we scan the ten N3 male carriers. The 'best' male that was produced from that generation was actually 96.15% recipient genome. Likewise, the 'best' male from the scan of the N_4 offspring was expected to be 98.53% recipient genome, based on the genome scan of fourteen carriers. The actual percentage genome of the 'best' male was 98.09%. There are many examples of 'best' males carrying a higher percentage of recipient genome than predicted. Conversely, on several occasions, we also observed less gain than predicted—for example, the N_5 C57Bl/Ks and the C3H/HeJ N_4. However, most of the estimated 'best' males and actual 'best' males have very similar percentage of recipient genome.

Additional observations can be made if we compare what we observed in Table 4 with the 'ideal' situation as predicted in Table 2. Males with a greater percentage of recipient genome than predicted by the ideal situation described in Table 2 are often observed, even though the number of male carriers scanned was often less than the ideal twenty. An extreme example of this is in the N_2 generation of the BALB/cByJ, in which only four male carriers were genome-scanned and yet the 'best' male observed had 86.36% recipient genome. Under 'ideal' conditions of scanning twenty male carriers, it was predicted that the 'best' male would have an 80.84% recipient genome (Table 2). The fact that we were able to advance the percentage of recipient genome so far by scanning only four male carriers reflects the fact that the 1.65-standard-deviation gain from twenty male carriers is the minimal gain that could be expected. The 1/0.25-n tail of the distribution can actually be a 1.65-SD gain or better; that is, the actual retention can appear anywhere in the tail of the distribution beyond the expected boundary of the standard deviation. None of the congenic strains in Table 4 reached the theoretical value at N_5 as presented in Table 2. In contrast to the data in Table 2, which do not account for keeping a target segment heterozygous, the data presented in Table 4 account for keeping the *Apoe* locus D/R. The number of markers used for the genome scan was, on average (depending upon the strain), 155. One of these is the *Apoe* locus. Therefore, 1/155 markers (0.66%) will be heterozygous, limiting the percentage of recipient genome to 99.7%. Both the C57Bl/Ks and the FVB/NJ reached this empirical maximum at N_5. The reason why the other strains did not quite achieve this percentage of recipient genome reflects that in no case did any of the strains truly follow the ideal number of twenty male carriers produced in each generation. However, it can be seen that in a true laboratory setting, where ideal theoretical numbers of mice cannot always be bred, five generations of backcrossing can reach the same, or better, percentage of recipient genome as N_8-N_9 of a traditional congenic-breeding scheme (Table 1). The global effects of this marker-assisted breeding regimen compared to the traditional methods are shown in cartoon form in Figure 1.

It should also be noted that the greatest gain of genome scanning is at the beginning of the backcrossing program. The magnitude of the standard deviations is exponentially greater in these earlier generations than in the final stages of congenic development (Table 2). Hence, although the genome scan of the N_2 generation is more labour-intensive than scanning of subsequent generations, having to cover the whole genome, it is also the most productive.

From the experience gained during this study, a few adaptations to the protocol outlined above are recommended. It is prudent to put both the 'best' and 'second best' male into breeding at each generation. This should help circumvent any potential problems of low viability. For example, the N_3 CAST 'best' male did not breed well and gave rise to only one male carrier. This severely reduced the probability of finding a male carrier that harboured the theoretically ideal percentage of recipient genome with which to carry on the subsequent generation. More acutely, a situation can arise whereby the chosen 'best' male does not breed at all, ceasing the continuation of the congenic strain. This was the case with the N_4 CAST. The problem of viability is not as acute during conventional congenic production because multiple matings are underway at each generation. If anything, high viability is probably inadvertently but fortuitously selected during conventional congenic production. A second advantage of mating both the 'best' and 'second best' males is that if twenty male carriers are not produced by the 'best' male, male carriers from the 'second best' male can be added to the pool for genome scanning. This simply attests to the fact that it is best to strive for twenty male carriers from a previously marker-selected parent, even if it is not the best parent. When we did this, a better subsequent-generation male was sometimes obtained from the 'second best' male than from the 'best' male. Obviously, the ideal situation of twenty male carriers (Table 2) from only the 'best' male of the previous generation is desirable, but laboratory conditions are often not ideal.

Conclusion

We have successfully used a marker-assisted congenic breeding strategy to move the *Apoe* null allele to multiple inbred strain genetic backgrounds. The goal of this study was to move a single specific locus to new genetic background. However, the method employed is equally pertinent to moving genomic regions. Indeed, more than one donor locus or region can be moved simultaneously (K.J.M., unpublished data). This marker-assisted breeding strategy can dramatically reduce the number of generations necessary to develop a congenic strain, and a congenic strain can now be established in 15–18 months. We have provided simple guidelines for the construction of congenic strains,

incorporating genotyping data with criteria for the selection of male carriers for subsequent breeding. An investigator can either calculate these values exactly with the formulas we have given or use the information in Tables 2 and 3 as a general rule of thumb. We now encourage other investigators to pursue this or similar methods of congenic production to introgress interesting loci into multiple genetic backgrounds in order to add to the already fertile resource of variant mouse strains pertinent to the study of complex human diseases.

Methods

Study overview. A null allele of the *Apoe* gene was generated (A. Plumb, personal communication) with the J1 ES-cell line, which is 129/Sv (+*Tyr* +*p*, *Ter*) derived. This *Apoe* null allele was then maintained on a 129/Sv (+*Tyr* +*p* , *Ter*) strain. We transferred this *Apoe* allele into inbred strains BALB/cByJ, C3H/HeJ, DBA/2J, CAST/Ei, C57BL/Ks and FVB/NJ by a breeding regimen commonly called 'speed congenic production'. Mouse breeding was carried out by the McLaughin Research Institute.

Genotyping. The differentiation of the *Apoe* null allele from that of the wild-type allele was determined under the following PCR conditions: 1.4 µl of ×10 PCR buffer, 1.12 µl of dNTP (2.5 mM), 0.15 U *Taq* polymerase, 0.16 µl (158 mM) of primer A (5′–CTCTGTGGGCCGTGCTGTTGGTCA-CATTGCTGACA–3′), 0.2 µl (126 mM) of primer B (5′–CTCGAGCT-GATCTGTCATCTCCGGCTCTCCC–3′) and 0.56 µl (45 mM) of primer C (5′–CGCCGCTCCCGATTCGCAGCGCATCGC–3′), 2.0 µl (20 ng) genomic DNA and 8.45 µl water; the total volume is 14 µl. Primers A and B identify the wild-type allele. Primers A and C identify the null allele.

A set of genetics markers was generated for the genome screen of each recipient strain. Markers were selected for screening from the MIT/Whitehead mouse map (Stein, L., Kruglyak, L., Slonim, D. & Lander, E. unpublished software, Whitehead Institute/MIT Center for Genome Research, 1995). Three or four markers were tested for polymorphisms between the strain 129/SvJ and the six recipient strains from each 10-cM interval across the genome. The DNA for all six recipient strains was obtained from the Jackson Laboratory. This analysis gave rise to approximately 150 usable polymorphic markers for each of the strain combination, resulting in a map density of roughly 10-cM intervals.

In the N_2 generation, a genome-wide set of markers at 20-cM intervals was used. For the N_3 generation, a second genome-wide set of markers was created that were interspersed between those markers used in the N_2 generation. For subsequent generations, markers were used only in regions that remained heterozygous (D/R) in the parent.

A list of the SSR markers used in this study and an indication of which are polymorphic between 129/SvJ and each of BALB/cByJ, C3H/HeJ, DBA/2J, CAST/Ei, C57BL/Ks and FVB/NJ are to be incorporated into the Mouse Genome Database at the Jackson Laboratory. In the interim, the authors will provide hard copies of the information upon request.

The PCR cocktail was 1.4 µl of ×10 PCR buffer, 1.12 µl of dNTP (2.5 µM), 0.2 U *Taq* polymerase, 1.05 µl of primers of each forward (6.6 µM) and reverse (6.6 µM), 1.0 µl (20 ng) genomic DNA and 8.35 µl water, the total volume being 14 µl. The PCR conditions were 94 °C for 2 min, 94 °C for 40 s, 55 °C for 50 s and 72 °C for 30 s, for 30 cycles. The products were run on a non-denaturing 8% acrylamide gel at 45 W, room temperature, for 3 h. After electrophoresis, the gels were stained with SYBR Green 1 and scanned on a fluorimager (Molecular Dynamics)[15].

Acknowledgements

We recognize E. Lander and G.M. Duyk for advice and helpful discussions. Thanks are extended to A. Plumb, who provided the starting 129/Sv (+Tyr +p, Ter) Apoe strain. We also thank N. Jenkins, N. Copeland and S.D.M. Brown for their helpful discussion and interest in this work and J. Meyer for her revision and comments on the manuscript. This research is supported by the Eli Lilly Corporation.

1. Markel, P.M., Bennett, B., Beeson, M., Gordon, L. & Johnson, T.E. Confirmation of quantitative trait loci for ethanol sensitivity. *Genome Res.* **7**, 92–99 (1997).
2. Oettgen, H.C. *et al.* Active anaphylaxis in IgE-deficient mice. *Nature* **370**, 367–370 (1994).
3. Ghosh, S. *et al.* Polygenic control of autoimmune diabetes in nonobese diabetic mice. *Nature Genet.* **4**, 404–409 (1993).
4. Taylor, B.A. & Phillips, SJ. Detection of obesity QTLs on mouse chromosomes 1 and 7 by selective DNA pooling. *Genomics* **34**, 389–398 (1996).
5. York, B., Lei, K. & West, D.B. Sensitivity to dietary obesity linked to a locus on chromosome 15 in a CAST/Ei×C57BL/6J intercross. *Mamm. Genome* **7**, 677–681 (1996).
6. Yui, M.A. *et al.* 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 (1996).
7. Morel, L., Yu, Y., Blenman, K.R., Caldwell, R.A. & Wakeland, E.K. Production of congenic mouse strains carrying genomic intervals containing SLE-susceptability genes derived from SLE-prone NZM2410. *Mamm. Genome* **7**, 335–339 (1996).
8. Green, E.L., ed. *Biology of the Laboratory Mouse* (McGraw-Hill, New York, 1966).
9. Silver, L. *Mouse Genetics, Concepts and Applications* (Oxford University Press, New York, 1995).
10. Plump, A.S. *et al.* Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-deficient mice created by homologous recombination in ES cells. *Cell* **71**, 343–353 (1992).
11. Paigen, B., Morrow, A., Brandon, C., Mitchell, D. & Holmes, P. Variation in susceptibility to atherosclerosis among inbred strains of mice. *Atherosclerosis* **57**, 65–73 (1985).
12. Zhang, S., Reddick, R.L., Piedrahita, J.A. & Maeda, N. Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E. *Science* **258**, 468–471 (1992).
13. Threadgill, D.W. *et al.* SSLPs to map genetic differences between the 129 strains and closed-colony, random-bred CD-1 mice. *Mamm. Genome* **8**, 441–442 (1997).
14. Simpson, E.M. *et al.* Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. *Nature Genet.* **16**, 19–27 (1997).
15. Kleyn, P.W. *et al.* Identification and characterization of the mouse obesity gene tubby: a member of a novel gene family. *Cell* **85**, 281–290 (1996).