

Research

# The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis

Robert W Williams<sup>\*†‡</sup>, Jing Gu<sup>\*‡</sup>, Shuhua Qi<sup>†‡</sup> and Lu Lu<sup>\*†‡</sup>

Addresses: <sup>\*</sup>Center of Genomics and Bioinformatics, <sup>†</sup>Center for Neuroscience, <sup>‡</sup>Department of Anatomy and Neurobiology, University of Tennessee Health Science Center, College of Medicine, Memphis, TN 38163, USA.

Correspondence: Robert W Williams. E-mail: rwilliam@nb.utm.edu

Published: 22 October 2001

*Genome Biology* 2001, **2(11)**:research0046.1-0046.18

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/2/11/research/0046>

© 2001 Williams et al., licensee BioMed Central Ltd  
(Print ISSN 1465-6906; Online ISSN 1465-6914)

Received: 25 June 2001

Revised: 31 July 2001

Accepted: 4 September 2001

A previous version of this manuscript was made available before peer review at <http://genomebiology.com/2001/2/6/preprint/0006/> (*Genome Biology* 2001, **2(6)**:preprint0006.1-0001.44)

## Abstract

**Background:** Recombinant inbred (RI) strains of mice are an important resource used to map and analyze complex traits. They have proved particularly effective in multidisciplinary genetic studies. Widespread use of RI strains has been hampered by their modest numbers and by the difficulty of combining results derived from different RI sets.

**Results:** We have increased the density of typed microsatellite markers two- to five-fold in each of several major RI sets that share C57BL/6 as a parental strain (AXB, BXA, BXD, BXH and CXB). A common set of 490 markers was genotyped in just over 100 RI strains. Genotypes of around 1,100 additional microsatellites in one or more RI sets were generated, collected and checked for errors. Consensus RI maps that integrate genotypes of approximately 1,600 microsatellite loci were assembled. The genomes of individual strains typically incorporate 45-55 recombination breakpoints. The collected RI set - termed the BXN set - contains approximately 5,000 breakpoints. The distribution of recombinations approximates a Poisson distribution and distances between breakpoints average about 0.5 centimorgans (cM). Locations of most breakpoints have been defined with a precision of < 2 cM. Genotypes deviate from Hardy-Weinberg equilibrium in only a small number of intervals.

**Conclusions:** Consensus maps derived from RI strains conform almost exactly to theoretical expectation and are close to the length predicted by the Haldane-Waddington equation ( $\times 3.6$  for a 2-3 cM interval between markers). Non-syntenic associations between different chromosomes introduce predictable distortions in quantitative trait locus (QTL) datasets that can be partly corrected using two-locus correlation matrices.

## Background

Recombinant inbred (RI) strains have been used extensively to map a wide range of Mendelian and quantitative traits [1]. They offer compelling advantages for mapping complex genetic traits, particularly those that have modest heritabilities. Each recombinant genome is replicated in the form of an entire isogenic line [2-6] and variance associated with

environmental factors and technical errors can be suppressed to low levels. This raises heritability and improves the prospects of mapping underlying quantitative trait loci (QTLs). We have recently used RI strains to map QTLs that generate variation in the architecture of the mouse central nervous system (CNS) [7-14]. The main advantage in this context is that the complex genetic and epigenetic

correlations between interconnected parts of the brain can be explored using complementary molecular, developmental, structural, pharmacological and behavioral techniques. Gene effects can also be tested under a spectrum of environmental perturbations and experimental conditions. RI strains can be exploited to expose gene-environment interactions and gene pleiotropy. These important facets of genetics can only be explored with difficulty using conventional mapping populations in which each genome is unique.

A third advantage of RI strains is that genotypes generated by different groups using a variety of methods can be pooled to generate high-density linkage maps. As a result, loci that segregate in RI sets can often be mapped with impressive precision without genotyping. This attribute was a significant advantage before the advent of efficient and easy PCR genotyping methods [15]. Unfortunately, over the past decade databases of RI genotypes have accumulated many typing errors. Each error expands distances between marker loci and degrades linkage, inevitably blurring associations between genotypes and phenotypes and making it difficult to map traits, whether they are Mendelian or quantitative in nature. The accumulation of false recombinations has become extreme in common RI sets. For example, the map of chromosome 1 in the complete BXD data set [16] is based on 160 linked marker loci and is an astonishing 1,305 cM long. This map is approximately 12 times the length of an F2 map of chromosome 1, and just over three times the length expected of an RI map of chromosome 1. The accumulation of typing errors has led to efforts to reconstitute maps using curated subsets of markers for which genotypes can be adequately and independently verified. Sampson and colleagues [17] assembled maps for the AXB and BXA strains that improved the utility of this set. Similarly, Taylor and colleagues [18] assembled comparable high-quality maps for the complete set of 36 BXD strains that are based almost entirely on easily typed and verified microsatellite markers.

Our aims complement this previous work. Our first aim has been to generate reliable high-resolution genetic maps for each of five widely used sets of RI strains: AXB, BXA, BXD, BXH and CXB. These RI sets all share C57BL/6 alleles, and they can be assembled into a BXN superset consisting of just over 100 lines. The introduction of the RI intercross (RIX) by Threadgill and colleagues [19,20] provides an impetus to precisely define recombination breakpoints in RI strains. RIX progeny are isogenic F1 hybrids made between pairs of RI strains. One hundred and one well-mapped RI strains could in principle be used to generate 5,050 well-defined isogenic and non-inbred RIX genotypes.

Our second aim has been to describe the recombination characteristics of typical RI strains and their chromosomes in a more theoretical context. We have empirically tested the Haldane-Waddington equation of map expansion in sib-mated RI strains. We have also tested relatedness among RI

lines, and measured deviations from Hardy-Weinberg equilibrium associated with 10-30 years of inbreeding, genetic drift, mutation and selection.

Our third aim has been to help resolve a serious but unrecognized problem in QTL mapping that arises from non-syntenic genetic correlations within mapping panels. Genetic correlations between intervals on different chromosomes can be high in RI sets and this can result in spurious results and false-positive QTLs. We provide detailed correlation matrices that can be used to detect and control for non-syntenic association.

Many of the files that form part of the analysis reported in this paper are available online (see Additional data files).

## Results

The results are divided into two sections. The first summarizes the RI consensus map and genotypes of individual strains. The second section considers the structure of the multi-generation meiotic recombination maps of RI strains. We highlight the problem of non-syntenic association that is a feature of RI genomes and we outline a solution that minimizes the risk of type-I and type-II error in QTL mapping studies.

### RI consensus maps of mouse chromosomes

Mapping complex genetic traits involves matching strain distribution patterns (SDPs) of genotypes with those of phenotypes. The utility of an RI set and the probability of successfully mapping any heritable quantitative trait or novel Mendelian trait is therefore a function of the number of well-defined and correctly positioned SDPs of marker loci. We therefore concentrated genotyping efforts on those intervals with comparatively low densities of fully typed microsatellite markers or those intervals that harbored large numbers of recombinations between neighboring markers. One goal in generating dense maps for each chromosome was to discover and verify as many recombination breakpoints and SDPs as possible using available microsatellite primer pairs. Ideally, in high-density genetic maps the number of markers should exceed the number of SDPs, and all recombination breakpoints in an set RI would be defined with subcentimorgan precision. We have worked with more than 1,600 microsatellite markers, a number that is still insufficient to reach this subcentimorgan goal. The density of markers on most chromosomes is, however, sufficient to locate the majority of recombination breakpoints within  $\pm 2$  cM.

Fewer than 25 common microsatellite markers had been typed on all major RI sets when we began this work. This number has been increased to 490 common markers (Table 1). These markers were used to assemble the consensus BXN maps - *B* for the C57BL/6 allele that all sets have in common and *N* for the not-B6 parental allele that differs among the four RI sets (A/J in AXB-BXA, DBA/2J in BXD, C3H/HeJ in

Table 1

## Summary of the numbers of microsatellite markers for which genotypes were generated or collected

Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	X
AXB-BXA	80	68	54	53	46	56	70	43	50	46	61	44	52	36	34	35	49	26	39	24
BXD	84	68	45	58	46	58	67	44	55	32	66	39	32	39	41	34	46	32	32	20
BXH	58	49	37	44	38	38	53	41	50	28	44	33	33	31	45	30	37	28	30	21
CXB	62	50	46	49	44	49	72	49	51	30	55	42	41	38	34	32	36	32	32	22
Common BXN	36	34	28	32	27	28	29	32	39	17	27	18	21	18	13	15	22	17	23	13
Total BXN	129	104	80	83	76	93	107	69	75	65	115	75	71	65	90	68	70	53	53	37

In the case of the AXB-BXA strains we pooled our genotypes with those generated by Sampson *et al.* [17]. Our new BXD data were pooled with genotypes of Taylor *et al.* [18], and CXB genotypes were pooled with the genotypes of Panoutsakopoulou and colleagues [55]. All data were eventually transferred to Map Manager QT and QTX. Both individual RI databases and the composite BXN database are available as text files formatted for use with Map Manager QTX files at [30]. The text files are compatible with Windows and Macintosh versions of Map Manager QTX and can be imported into a text editor or spreadsheet program.

BXH, and BALB/cByJ in CXB). The set of 490 shared markers are supported by an additional 1,089 MIT markers that we or other groups have typed in at least one RI set (Figure 1). In the BXN database summarized in Table 1, any pair of RI sets shares between 500 and 600 fully genotyped markers. The two largest RI sets, AXB-BXA and BXD, have been typed at 591 common markers. The composite BXN maps are based on a total of just under 1,600 microsatellite makers and just over 100 RI strains (Table 1, Figure 1).

#### Undiscovered recombinations and SDPs

The number of recombinations in RI sets still significantly exceeds the number of SDPs that have been unequivocally defined. On the basis of current marker density we estimate that we have defined from 37% (AXB-BXA) to 59% (CXB) of the total set of SDPs (Table 2). The entire BXN set contains approximately 4,800 known recombination breakpoints (Tables 2,3). There are likely to be another 400 breakpoints that we have not yet detected. To discover 623 (41%) of the 1,492 SDPs in the BXD set required 936 selected markers. Recovering the majority of the remaining SDPs could require an additional 1,000 to 1,500 well placed marker loci. The density of informative microsatellite markers is not yet sufficient to define many more SDPs in the BXN set, but once SNP and microsatellite maps have been fully integrated into chromosome sequence databases, it will be straightforward to generate additional markers and use these to define all 5,000-6,000 SDPs in the BXN set (see [21] for integrated MIT and Roche SNP data files).

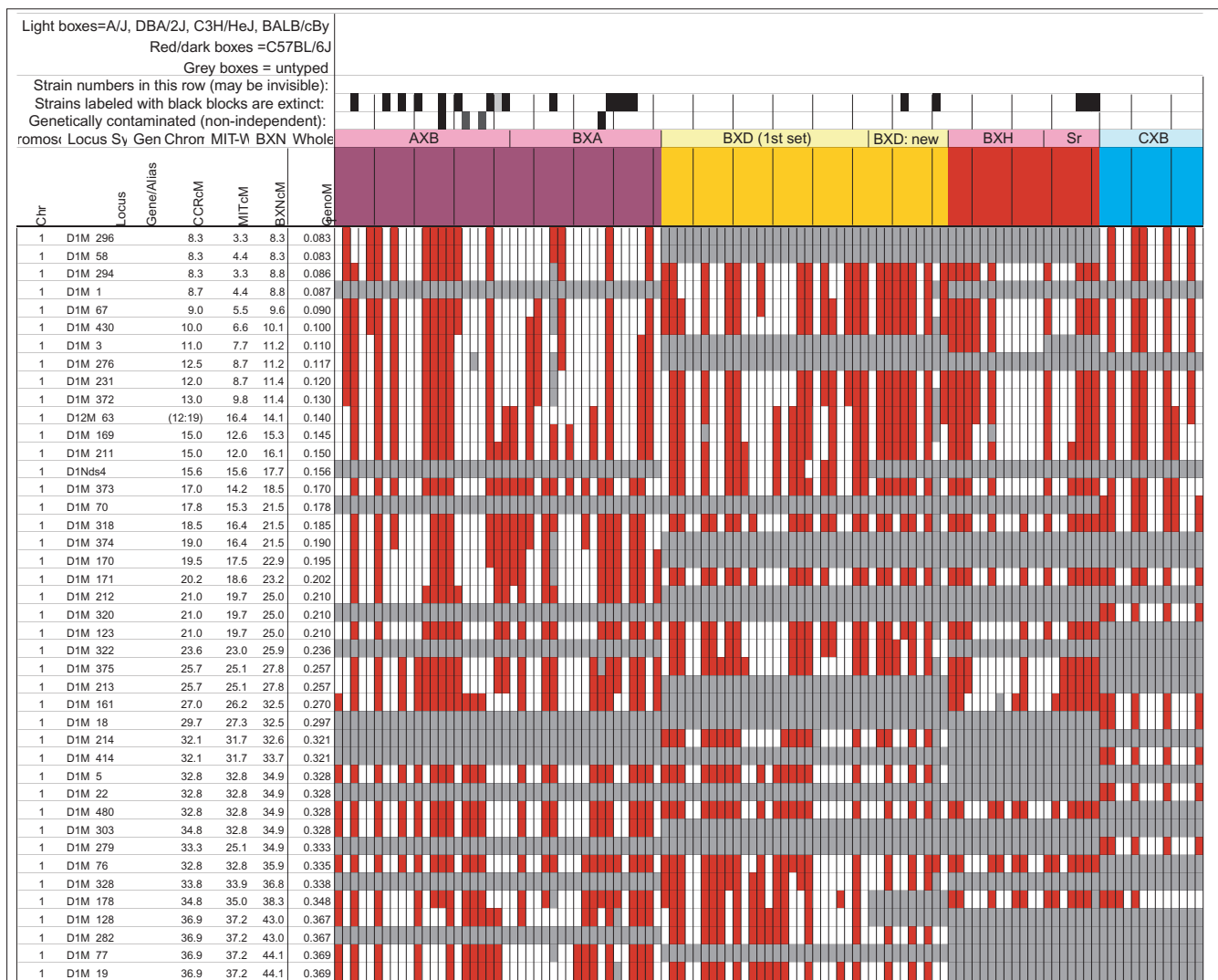
#### Error checking

To minimize genotyping errors we retyped many markers, particularly those that were associated with unusually large numbers of recombination events. We were particularly interested in minimizing the number of genotypes that appeared to be associated with two closely located recombination events - what are sometimes referred to as double-recombinant haplotypes. These haplotypes appear to be the

result of two separate crossover events, one of which is just proximal to a particular marker and the other just distal to the same marker. For example, the haplotype of a short chromosome interval *-BBBNNBB-* is associated with two recombinations that flank the central marker with the *N* genotype. Because of interference, the occurrence of two recombinations within 10 cM is highly improbable in an F2 intercross, and consequently, double recombinants are often used as a measure of genotyping error or incorrect marker order. In RI strains, however, recombination events accumulate over many generations, and two or more recombinations can therefore be extremely close to each other and can produce true double-recombinant haplotypes. It is therefore necessary to verify, rather than discard, all apparent double recombinants in RI strains. We checked our own marker genotypes and the majority of microsatellite markers typed by other investigators for whether they were associated with double recombination events in one or more RI strains. When two or more strains contributed to double recombinants, we usually retyped all strains. Approximately 150 double-recombinant haplotypes (and 300 false recombinations) were eliminated in the process of error checking. Our genotypes therefore differ from those of many microsatellites reported in original publications and listed in the Mouse Genome Informatics Release 2.5 [16]. In a few instances, our revisions have generated new (but verified) double-recombinant haplotypes.

We discovered unexpected polymorphisms at several loci in a few lines and all were scored as unknown (U) (Table 4). The clustering of aberrant products in AXB13 and AXB14 is consistent with the common origin of these strains from a partly inbred progenitor line. However, the genotypes of the other three sets of strains (for example, AXB1 and AXB3) are generally completely independent.

PCR primer pairs in several intervals gave two bands consistent with a genuine heterozygous genotype. Heterozygous

**Figure 1**

The BXN map of the mouse genome. The full data table is available in several formats (graphic, text, and Map Manager QTX) as Additional data files and at [30]. Column definitions from left to right: Chr, chromosome assignment based on BXN data set. Our assignments differ in a number of cases from those of the Chromosome Committees' Reports. Locus, an abbreviated version of the locus symbol. To improve legibility we have truncated D1MitNN to D1M NN. CCR cM, the position of the locus given in the most recent chromosome committee reports (2000 or 2001). MIT, the position of the locus given in databases at the Whitehead Institute. BXN, position computed from the current RI data set adjusted for map expansion. GenoM, whole-genome position in morgans with a 5 cM buffer (0.05 M) between chromosomes. This GenoM column can be used to construct whole-genome LOD score plots.

loci are rare among fully inbred RI strains but they are fairly common among new BXH strains that were genotyped at the 10th to 16th generation of inbreeding. In scoring recombination frequency we treated all heterozygous loci and intervals as if they had not been typed. Mutations in microsatellite loci may be responsible for some heterozygosity [22].

#### Changed locus order

The order of loci of the BXN consensus maps generally conforms to that of the chromosome committee reports (CCR) and the MIT-Whitehead genetic maps (Figure 1). In about

130 instances we have changed the order of loci over short intervals. For example, *D1Mit276* and *D1Mit231* on proximal chromosome 1 do not recombine in the MIT F2 cross, but in the BXN set there is a single recombination between these markers in BXA11 that is most consistent with a reversal of order relative to the CCR (compare the columns labeled CCRcM, MITcM, and BXNcM in Figure 1). The only non-trivial discrepancy was on proximal chromosome 15. We reordered approximately 32 loci on chromosome 15 to improve linkage statistics. We have not attempted to integrate the BXN data with numerous other mapping panels,

**Table 2**

**Comparison of recombination characteristics of RI sets**

RI set	N strains*	N typed†	Total markers	New markers‡	Total SDPs	Total R	% SDP	DR	% DR	R per strain
AXB-BXA	37 (28)	37 (41)	965	577	669	1775	37.3%	114	12.8%	43.3
BXD	35 (34)	36	936	551	623	1492	41.2%	95	12.7%	41.4
BXH	16 (16)	19	768	739	436	848	50.2%	82	19.3%	44.6
CXB	13 (13)	13	867	687	406	671	58.8%	63	18.8%	51.6
BXN	104 (91)	109	1,575	1,575	1,220§	4,786	25.4%	354	14.8%	43.9

\*Total number of genetically independent strains per RI set genotyped in this study. The numbers in parentheses are numbers of strains available as live stock from the Jackson Laboratory. †Total number of strains per RI set genotyped in all studies. We genotyped 107 of these 109 strains (BXA9 and BXD41 were not genotyped). Four of the AXB-BXA strains are not genetically independent and the effective number of genotyped strains is actually 37 rather than 41. ‡This number includes both newly genotyped markers and markers for which genotypes were changed as a result of retyping. §Due to overlap of SDPs this number is not a sum of the overlying four values. The total number of unique SDPs was computed from the composite BXN set directly. Total SDPs is the total number of unique strain distribution patterns identified in each set. Total R is our estimate of the total number of independent recombinations (R) represented in each complete set of RI strains, including members of RI sets that are now extinct. This number will underestimate true numbers of recombination breakpoints by about 7.5% due to the limited number of available markers on some chromosome regions. % SDP is our estimate of the percentage of the total number of SDPs that have actually been identified in association with particular markers. DR is the sum of recombinations associated with double recombinant haplotypes (see text) in each RI set. % DR is the percentage of recombinations associated with these double recombinant haplotypes. R per strain is the estimate of the average number of recombination events accumulated by individual RI strains.

**Table 3**

**Recombinations per chromosome**

Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	X
AXB-BXA	132	128	109	101	117	87	98	75	75	103	96	82	84	70	75	88	82	58	61	54
BXD	111	128	79	88	87	89	89	65	62	52	88	78	65	48	60	68	68	71	58	38
BXH	65	57	46	65	40	40	37	48	27	29	61	33	41	42	48	42	31	36	36	24
CXB	36	44	41	47	25	36	37	39	29	47	37	32	30	29	31	23	30	34	29	15
BXN	344	357	275	301	269	252	261	227	193	231	282	225	220	189	214	221	211	199	123	131
BXN cM*	344	357	275	301	269	252	261	227	193	231	282	225	220	189	214	221	211	199	123	131
CCR cM*	104	108	86	83	94	74	72	74	64	68	77	59	71	63	60	68	56	56	56	72
Expansion	3.3	3.3	3.2	3.6	2.9	3.4	3.6	3.1	3.0	3.4	3.7	3.8	3.1	3.0	3.6	3.3	3.8	3.6	2.2	1.8

\*The distance in centimorgans between the most proximal and the most distal markers on each chromosome. The mean number of strains typed at each marker is approximately 100 and thus distances in centimorgans match the actual number of recombination events per chromosome. In the case of the CCR maps we have truncated map lengths to match the most proximal and distal markers genotyped in the BXN set.

and it is likely that original CCR order will often be well supported by other large mapping panels or rapidly improving physical maps. Full sequence data will soon resolve these minor inconsistencies.

**Reassigned microsatellite loci**

A number of microsatellite loci were reassigned to locations on chromosomes other than those expected on the basis of their original assignments (Table 5). Mapping data in one or more of the RI sets is consistent with a reassignment of 16 microsatellite loci to different chromosomes. All of these reassignments are provisional, particularly those with LOD scores of less than 10. In several cases (for example, *D10Nds10*) we have reassigned microsatellite loci typed by other investigators that now are linked to new and firmly mapped markers. All primers used to amplify these

microsatellites (except *D10Nds10*) were resynthesized to confirm that they are identical to those originally specified by Dietrich and colleagues [23].

Individual maps are based on genotypes of as few as 37 markers (chromosome X) to as many as 129 markers (chromosome 1) per chromosome (Table 1). The mean separation between markers is approximately 1 cM (0.95 cM using CCR maps as a reference and 0.87 cM using the RI maps themselves). When the 577 markers that do not have unique SDPs are excluded from the analysis, the average separation increases to 1.2 cM using CCR maps and 1.4 cM using the RI data. Typical resolution of the BXN set for mapping a Mendelian trait is 1-2 cM. Approximately 90% of the mouse genome is currently less than 2 cM from a typed microsatellite marker in the RI set. The asymptotic resolution of the set

**Table 4**

Novel or unexpected PCR products of microsatellite loci					
Marker	cM	Strains	N*	B*	Observed*
D6Mit61	53.0	AXB13, AXB14	146	136	142
D6Mit116	6.0	AXB13, AXB14	114	123	108
D15Mit175	6.7	AXB1, AXB3	164	178	140
D6Mit264	3.2	CXB6, CXB12	116	124	120
D9Mit162	28.5	BXH2, BXH3, BXH6, BXH8	122	140	118

\*Product length ( $\pm 2$  bp) of PCR products generated with standard MIT primer pairs. N = not-B6 allele (A/J or BALB/cBy), or C3H/HeJ), B = C57BL/6 allele.

of BXN strains given infinitely dense maps in which every possible SDP has been characterized would average about 0.3-0.4 cM. There are currently 14 poorly typed regions. These regions are operationally defined as intervals of 5 to 12 cM between adjacent markers (Figure 2). The largest is on proximal chromosome 2 between 9 and 21 cM (Figure 1).

#### Strain independence

Several RI strains share common haplotypes and recombination breakpoints. This non-independence of RI lines will distort genetic maps. To systematically search for and eliminate partial duplicate RI lines we constructed a genotype similarity matrix for all strains using the QTL analysis program Qgene [24,25]. An example of a small part of this matrix is illustrated in Table 6 for the CXB set.

As already noted by Sampson *et al.* [17], three sets of AXB and BXA strains show high genetic similarity, and genotypes of four strains should be excluded from most genome-wide mapping panels. Phenotype data obtained from members of the three groups listed below should often be collapsed and treated as a single strain.

Group 1 consists of BXA8 and BXA17, which have 99.8% genetic identity. Only two markers are known to be polymorphic, *D3Mit392* and *D6Mit108*. The polymorphism at *D6Mit108* has been verified using independent DNA samples from these two strains. BXA17 is actually a direct derivative of BXA8 separated in 1996-97 [17]. Any divergence in genotypes or phenotypes is due to the recent generation and fixation of new mutations in these two separately maintained lines. Group 2 comprises AXB18, AXB19, and AXB20. There is 97 to 99% identity among any of the three pairs. Group 3 comprises AXB13 and AXB14, which have 92% identity. These three sets of strains were treated as three single strains when analyzing recombination frequencies.

The mean allele similarity of the remaining strains averages almost exactly 50%. The distribution of values is symmetrical

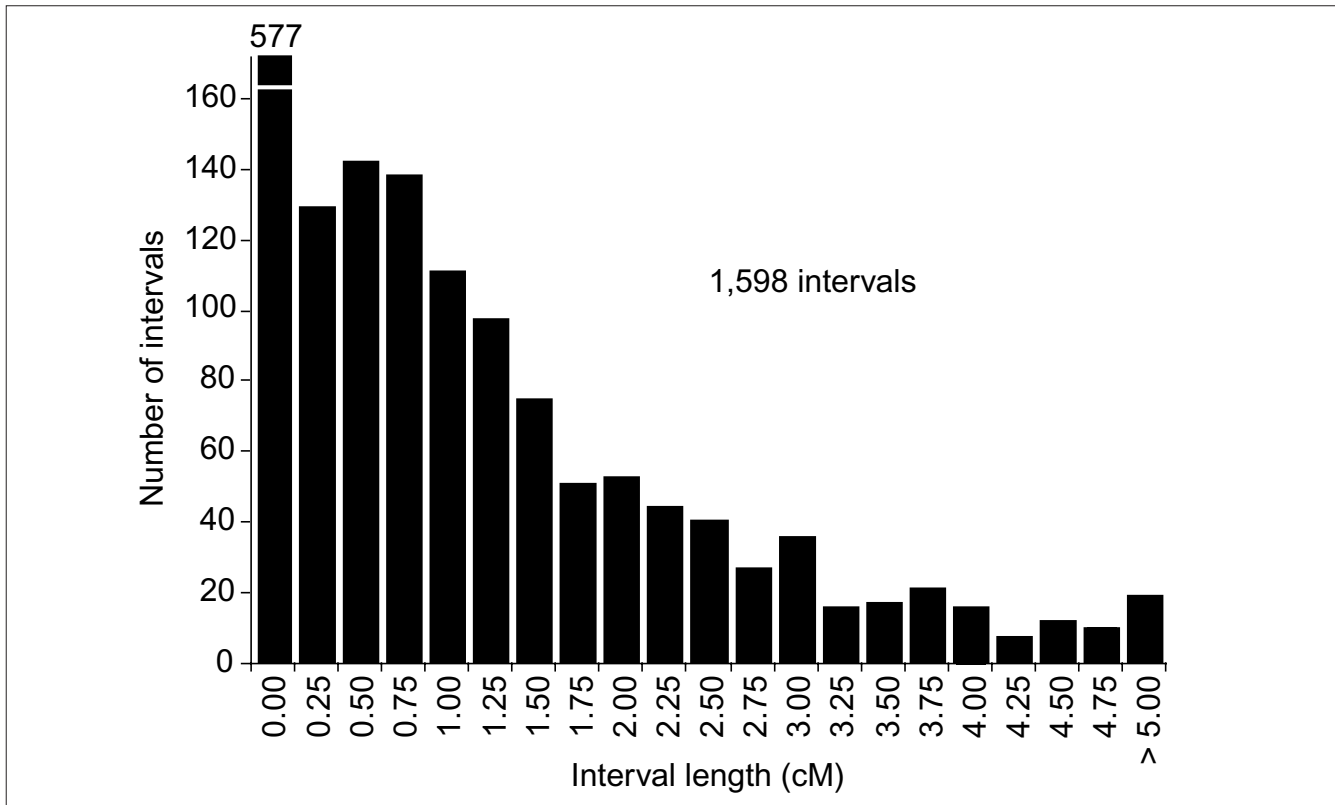
**Table 5**

Loci mapped to unexpected chromosomes						
Symbol	New chromosome	cM	LOD	Linked to	cM	Sets
D12Mit63	1	13.5	22.9	D1Mit169	14.0	BXN
D15Mit139	3	84.9	32.3	D3Mit116	84.9	BXN
D1Mit167	5	0.5	3.9	D5Mit346	0.5	CXB
D7Mit284	6	42.0	3.9	D6Mit230	43.0	CXB
D12Mit38	7	49.5	29.7	D7Mit38	49.8	BXN
D10Nds10	8	44.6	7.5	D8Mit266	44.5	BXD
D10Mit198	9	28.6	26.3	D9Mit4	29.0	BXN
D8Mit18	11	56.0	3.9	D11Mit98	58.0	CXB
D13Mit217	12	11.5	10.0	D12Mit106	12.0	AXB-BXA
D1Mit464	12	13.0	32.2	D12Mit136	13.0	BXD, BXH, CXB
D1Mit163	13	48.0	3.9	D13Mit107	3.0	CXB
D18Mit128	14	46.9	5.8	D14Mit265	48.0	BXD
D15Mit19	17	3.5	16.8	D17Mit267	3.0	BXN
D14Mit207	19	21.0	7.2	D19Mit13	20.0	BXD, BXH
D4Mit50	19	55.5	21.1	D19Mit6	55.7	AXB-BXA
D6Mit324	X	26.5	28.3	DxMit1	27.0	BXN

about the mean (Figure 3) with the great majority of strain pairs falling in the range of 30-70% similarity. The highest remaining similarities within RI sets are between BXD13 and BXD41 (74%), AXB6 and AXB17 (73%), BXH2 and BXH9 (71%), AXB6 and AXB12 (70%), BXD28 and BXD33 (69%), BXD19 and BXD29 (68%) and AXB11 and AXB14 (67%). These values are not significantly higher than the similarity scores typically noted across RI sets.

#### Residual heterozygosity

In theory, a set of 75,000 genotypes generated across the genome of 100 RI strains should detect only a single residual heterozygous loci at generation F55 of inbreeding (Figure 3, lowest line; the inbreeding coefficient at F55 is 0.99998812). DNA from most lines was extracted in the 1990s at F generations between F20 and F70 (see Materials and methods). We detected a total of 13 strains that were still heterozygous (BXA20 from *D1Mit77* to *D1Mit490*; AXB21 from *D2Mit102* to *D2Mit420*, AXB24 at *D3Mit62*, BXA23 at *D5Mit95*, AXB3 and BXA16 at *D12Mit167*, BXA20 from *D13Mit224* to *D13Mit254*; BXD31 at *D9Mit243*, BXD34 at *D7Mit281*, BXD37 at *D1Mit83*; BXH12 at *D1Mit417*, BXH10 at *D12Mit167*; CXB8 from *D1Mit361* to *D1Mit291*). DNA samples were taken from single animals of each strain and for this reason these estimates of residual heterozygosity underestimate the total heterozygosity about two-fold.



**Figure 2**  
Histogram of interval length in centimorgans between neighboring microsatellite markers in the BXN set.

The central part of chromosome 1 is interesting because it is heterozygous in three strains (BXD37, BXH12 and BXA20). There is also an interval of approximately 2.5 cM that is apparently maintained in heterozygosity in AXB21 on chromosome 2. Such maintenance should be accompanied by reduced fecundity in this line if homozygotes are lethal or sublethal. This would account for poor breeding performance. It is also possible that the heterozygosity is the result of a mutation, but if this were the case we would expect novel length polymorphisms, and the two alleles were usually the expected parental lengths.

**Structure of RI genomes**

*RI mean map lengths*

The mean frequency of recombinations,  $C_{RI}$ , between two linked markers in an RI strain generated by breeding siblings is approximately  $4c/(1 + 6c)$  where  $c$  is the recombination fraction per meiosis [26,27]. An infinitely dense RI map should average four times the length of the conventional one-generation F2 map. Most expansion is achieved in the first few generations, and by F7 the genetic map is approximately three times the length of an F2 map (Figure 4). The expectation is that a map based on loci that are spaced at intervals of 1 cM ( $c = 0.01$  in an intercross) will be expanded approximately 3.66-fold. Similarly, a low-density map based

on markers that are spaced at 16 cM intervals will be expanded two-fold. F2 and N2 maps generated using uniform typing procedures typically have a cumulative length of 1,300 to 1,400 cM. Five conventional crosses that we generated (four F2s and one N2, each genotyped at 91 to 148 loci) average  $1,320 \pm 50$  (standard error of the mean) cM in length. In comparison, the fully error-checked native BXN map is approximately 3.6- to 3.7-fold longer, or a total of 4,786 cM. The expansion averages approximately 3.4-fold when the comparison is made to the CCR consensus maps (Figure 5, Table 3). The expansion between common proximal and distal markers ranges from 2.8 in chromosome 5 to 3.8 in chromosome 12. In general, the expansion estimate of 3.6-fold agrees well with the Haldane-Waddington expectation, given a mean spacing between neighboring markers of 2-3 cM. The X chromosome only recombines with half the frequency of the autosomes, and for this reason its expansion is only 1.8-fold.

*Comparison with other maps*

The summed length of all chromosomes is approximately 1,413 cM when values are converted from RI recombination frequencies to those expected of typical single-generation meiotic maps. The corresponding CCR maps have a cumulative length of 1,494 cM between the same markers.

**Table 6****Sample of the strain similarity matrix**

Strain	CXB13	CXB12	CXB11	CXB10	CXB9	CXB8	CXB7	CXB6	CXB5	CXB4	CXB3	CXB2
CXB12	0.55											
CXB11	0.44	0.42										
CXB10	0.57	0.53	0.40									
CXB9	0.35	0.47	0.53	0.50								
CXB8	0.52	0.54	0.59	0.51	0.50							
CXB7	0.53	0.52	0.53	0.43	0.46	0.67						
CXB6	0.53	0.54	0.50	0.49	0.45	0.49	0.53					
CXB5	0.51	0.37	0.53	0.47	0.43	0.47	0.46	0.50				
CXB4	0.51	0.61	0.52	0.52	0.43	0.49	0.48	0.54	0.48			
CXB3	0.47	0.46	0.52	0.51	0.49	0.53	0.49	0.45	0.51	0.49		
CXB2	0.58	0.53	0.51	0.40	0.54	0.53	0.52	0.45	0.56	0.45	0.48	
CXB1	0.48	0.44	0.51	0.48	0.51	0.42	0.53	0.39	0.43	0.50	0.47	0.43

The fraction of identical genotypes was computed for all two-way combinations of 109 RI strains. Those pairs of strains for which the percentage of shared genotypes was greater than 75% were flagged and one member of the pair was eliminated from the BXN set. Corresponding matrices for AXB-BXA, BXD, BXH and the complete BXN matrix are available online at [30] in text format.

The MIT-Whitehead microsatellite maps have a cumulative length of approximately 1,384 cM. The agreement is excellent.

*Recombination density per RI strain*

Individual RI strains contain an average of 47 recombinations with a range that typically lies between 40 and 60 (Figure 5). The 13 CXB strains are associated with a total of 671 recombinations, an average of 52 per strain. The BXD strains are associated with approximately 1,500 recombinations, an average of about 42 per strain, and approximately one recombination per centimorgan on a standard genetic map (Tables 2,3). There is considerable variation in the total load of recombinations and map expansion per strain: from a low expansion of 2.24 in BXD40 (the RI strain with the fewest recombinations) to a high expansion of about 6 in BXH6 (Figure 5). These estimates are systematically deflated by a failure to discover recombinations in sparsely mapped regions (regions where the recombination fraction  $c$  is as high as 0.1) but are inflated by residual typing errors and errors of marker order.

*Recombination density per chromosome*

Single chromosomes in RI strains accumulate as many as 12 recombinations, but across the whole set the recombination density averages about 2.4 recombinations per chromosome. The mean extends from 3.47 recombinations for chromosome 1 to 1.88 for chromosome 9. A Poisson model fits the distribution of recombination events per chromosome reasonably well and most chromosomes have insignificant chi-square ( $\chi^2$ ) values (Figure 6). High  $\chi^2$  for individual chromosomes are generally due to a small number of apparently highly recombinant chromosomes in particular strains.

These highly recombinant chromosomes are probably associated with residual typing errors or incorrect marker order.

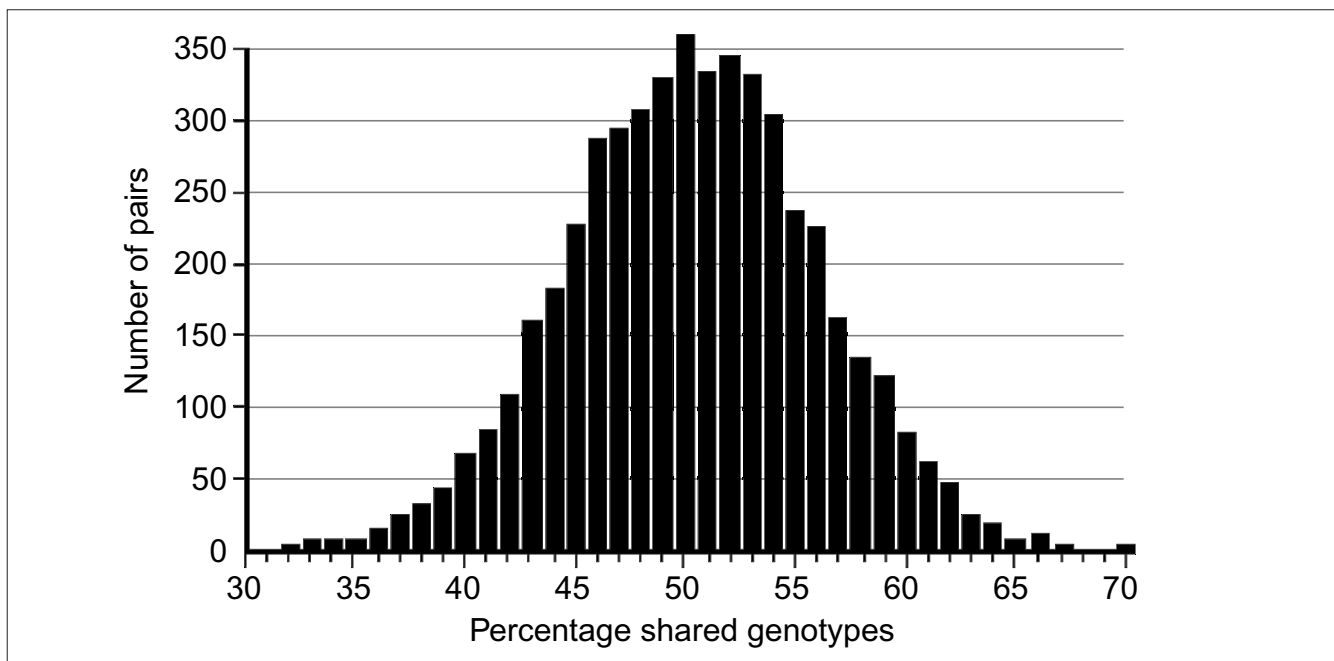
*Segregation distortion and Hardy-Weinberg equilibrium expectation of allele fixation in RI sets*

In the absence of selection, approximately 50% of the strains should have inherited  $B$  alleles at each marker. A  $\chi^2$  statistic can be used to assess whether the segregation ratio of a particular marker differs significantly from expectation. Only the 11 intervals listed in Table 7 have  $\chi^2$  values that are significant at the 0.01 level. Eight of 11 intervals are biased in favor of  $B$  alleles. This is most extreme on chromosomes 1, 15 and X, where there are about twice as many strains with  $B$  alleles as  $N$  alleles. The opposite pattern is seen on chromosomes 9, 11 and 12. Given the large number of comparisons, many instances of segregation distortion may be type-I statistical errors. In collaboration with the Mammalian Genotyping Service [28], we recently genotyped a tenth-generation advanced intercross between C57BL/6J and DBA/2J (genotype data for this cross is available at [29]). It is therefore possible to test whether similar segregation distortion patterns are present in this related multi-generation cross. The short answer is that the segregation distortions noted in the BXN RI strains are replicated in 6 of 11 intervals. The correlation between ratios of alleles (logarithm of  $B:N$ ) in these intervals was positive ( $r = 0.41$ ). It is therefore likely that several of the intervals marked in Table 7 with asterisks represent regions that harbor loci that affect fitness.

*Non-syntenic associations*

One important issue in using RI strains for mapping complex traits is that intervals on different chromosomes





**Figure 3**

Genetic similarity of RI strains. The percentage of identical genotypes was computed for all two-way combinations of 108 RI strains. Those pairs of strains for which the percentage of shared genotypes was greater than 75% (see text) were flagged and one member of the pair was eliminated from the BXN set.

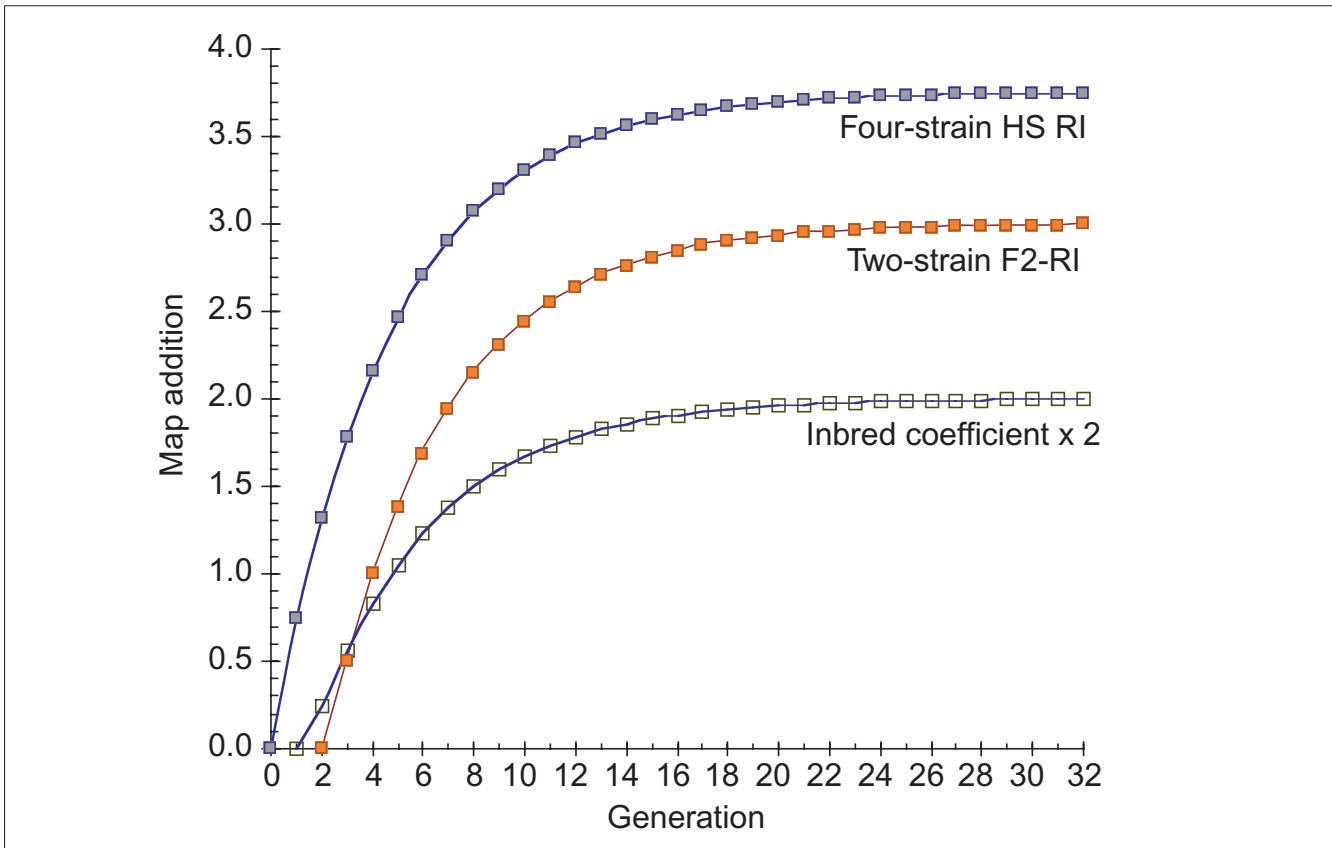
can become tightly associated in a statistical sense. This non-syntenic association can arise either as a result of random fixation of alleles on different chromosomes during the production of RI strains or can arise as a result of selection for particular combinations of alleles on different chromosomes. Similar patterns of non-syntenic disequilibrium are common in recently admixed human populations and often lead to false-positive signals when mapping complex traits. In mice, even a modest selection coefficient expressed over ten generations of inbreeding can generate positive and negative non-syntenic disequilibrium throughout the genome. For example, if the combination of *B* alleles on distal chromosome 1 and *B* alleles on proximal chromosome 19 is favorable for fitness, then these two intervals will effectively be in linkage disequilibrium in the final RI set. Disequilibrium can also take the form of strong negative correlations and *B* alleles may be associated strongly with the group of *N* alleles.

We searched for marked deviations from the expected Hardy-Weinberg two-locus equilibrium by making a series of large correlation matrices of SDPs of marker pairs (see [30] for a variety of correlation matrices). This was done for the entire BXN set and for the constituent RI sets. Figure 7 summarizes the most extreme positive and negative correlations among the composite set of 102 independent BXN RI strains. Whether due to chance fixation, selection or epistasis, non-syntenic associations of the sort illustrated in

Figure 7 are a major source of both false-positive and false-negative results in using RI sets for mapping. It is helpful to examine the correlation matrix once a set of QTLs has been provisionally mapped to see how summed effects of single or multiple QTLs might produce spurious QTLs in regions not actually associated with trait variance.

#### *Controlling for non-syntenic association*

Non-syntenic associations among loci and intervals can be computed in advance of QTL mapping. It is therefore possible to statistically control for genetic correlations that are built into different RI sets. For example, in Figure 7 the genotypes at marker *D1Mit83* can be partly predicted by genotypes at markers on chromosome 7 and chromosome 10. If the genotype at *D1Mit83* is treated statistically as a dependent variable and markers on chromosomes 7 and 10 are used as predictors, then one can compute the residual genotype, or independent contribution of *D1Mit83* and any other marker or interval to the quantitative trait. Unlike composite interval mapping, the set of controlled loci will necessarily vary for each marker and interval. This procedure will reduce type I error but will produce a regional loss of statistical power. The correction will introduce blind spots in a genome scan. In extreme cases (usually small RI sets), intervals that can be perfectly predicted by small numbers of other non-syntenic intervals will effectively be eliminated from a mapping study and QTLs in those intervals will be missed. For this reason, it is essential to perform each a



**Figure 4**  
 Progressive expansion of RI genetic maps during inbreeding. The middle series of points (red) that start at generation 2 shows the addition of map length - and the proportional increase in the numbers of recombination breakpoints - relative to a standard one meiotic generation F2 map. For example, at generation 7, approximately two map lengths have been added to the initial map. By F24 the total RI map is almost precisely four times as long as a standard F2 map. This same addition characterizes other diallele crosses that start near Hardy-Weinberg equilibrium, including advanced intercrosses. A two-strain G8 advanced intercross with a 6,000 cM map length would ultimately produce a G8 RI set with map length of  $6,000 + 3 \times 1,400$  cM = 10,200 cM. The upper series of points (blue) illustrates the accumulation in map length in a four-strain intercross at Hardy-Weinberg equilibrium at generation 0. This cross will gain up to 3.75 map equivalents. The lowest set of points is the inbreeding coefficient at each generation. For a tabulation of these data and methods for calculating two- and four-strain expansion values see [30].

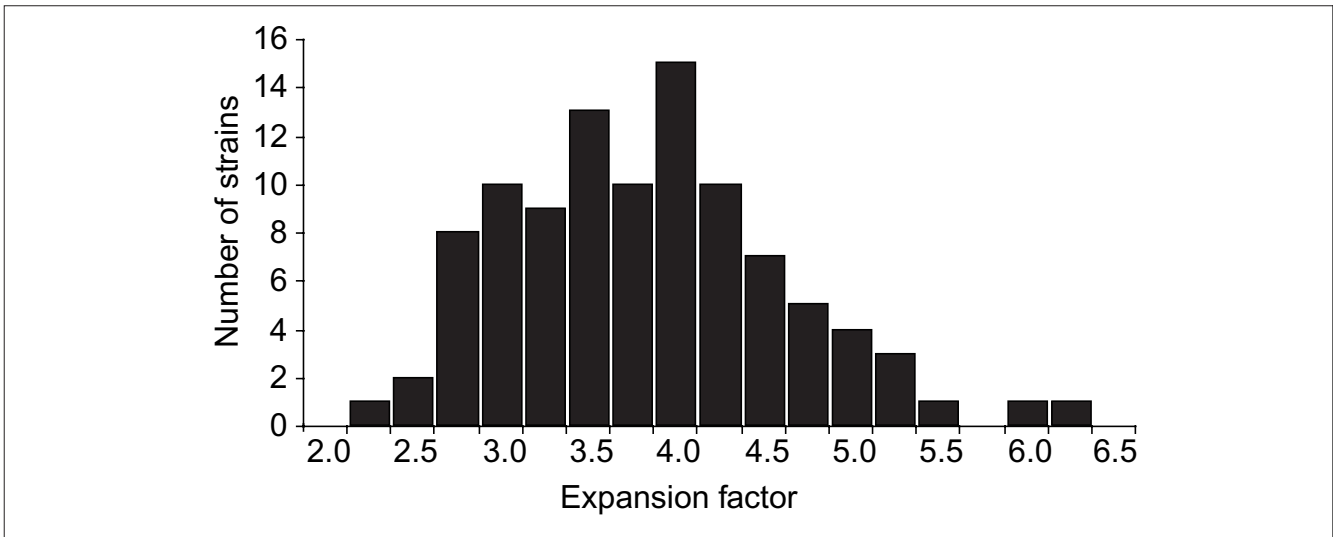
genome-wide scan both with and without control for non-syntenic association. Single QTLs may occasionally be assigned to two or more physically unlinked intervals.

**Discussion**

Recombinant inbred strains are currently one of the best genetic resources for exploring phenotypic variance modulated by complex mixtures of genetic and environmental factors. A renewable resource of genetically defined genomes is an important advantage in exploring gene pleiotropy, genetic correlation and reaction norms [1-3,7,8,10-12,27]. For example, Eleftheriou and colleagues [31] exploited the CXB set to test effects of subtle environment differences (animals reared in Italy or at the Jackson Laboratory) on brain weight, and we have been able to revisit this same phenotype in the CXB set after an interval of 25 years. With the improved set of

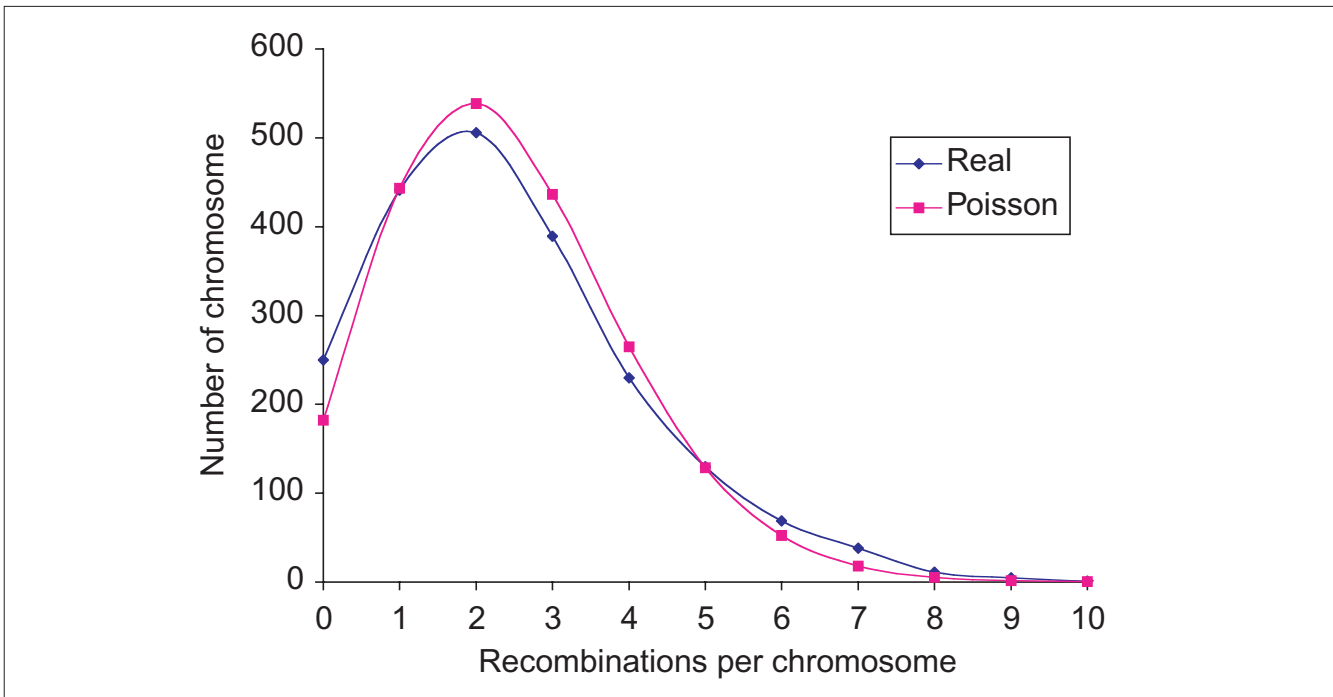
fully typed markers it is now feasible to map sets of QTLs under different environmental conditions, including temperature, pathogen load and food source, using RI strains. The modest number of RI strains, among other considerations, has, however, hindered their widespread adoption by mammalian geneticists. To improve the utility and power of complex trait analysis and to provide a better basis for collaborative QTL mapping, we have increased marker density in several of the major sets of RI lines and have merged data from over 100 mouse RI strains using a framework based on 490 shared markers. Approximately 1,000 unique SDPs (an average of about one per 1.5 cM) have been defined and mapped in the collected set. Three to four times as many SDPs remain to be discovered in the BXN set.

At the current marker density the cumulative RI map is about 5,000 cM long, roughly 3.6 times the length of standard



**Figure 5**

Mean expansion of the genetic map in RI strains. The average is approximately 3.7 for 100 independent RI lines. The x-axis can also be considered as the mean number of recombinations per 100 cM in different RI strains. This can be transformed into the total number of recombinations per strain by multiplying by the genetic length of the mouse genome in morgans (approximately 14 morgans; 2.25x = 31.5 recombinations/strain, 3x = 42 recombinations/strain, 4x = 56 recombinations/strain; and 6x = 84 recombinations/strain).



**Figure 6**

Density of recombinations for all autosomes compared to a Poisson model. We scored the number of recombinations for each of 2,072 chromosomes (all strains; chromosome X excluded). The mean number is 2.43 recombination breakpoints per chromosome. The particular distribution assumes all 19 autosomes have a length of about 70 cM and this simplification accounts for the high  $\chi^2$  (125,  $p < 0.001$ , 10 df). Of 250 non-recombinant chromosomes observed only 182 were expected. There are also significantly more chromosomes with an apparent excess of recombinations. These deviations are of course expected because short chromosomes (< 70 cM) will contribute more non-recombinants and long chromosomes (> 70 cM) will contribute more highly recombinant chromosomes than predicted by the model.

**Table 7****Hardy-Weinberg deviations in the BXN**

Proximal marker	cM	Distal marker	cM	Genotypes		BXN L(B:N)	AI L(B:N)
				BB	NN		
D1Mit14	75	D1Mit455	92	71	38	0.27	0.54*
D2Mit343	86	D2Mit343	86	69	40	0.24	-0.12
D3Mit28	45	D3Mit28	45	66	39	0.23	-0.14
D9Mit4	29	D9Mit289	38	38	71	-0.27	-0.41*
D10Mit42	44	D10Mit42	44	68	41	0.22	0.43*
D11Mit2	2	D11Mit296	10	38	71	-0.27	0.02
D12Mit110	22	D12Mit89	24	37	71	-0.28	0.00
D15Mit180	14	D15Mit105	42	71	36	0.29	-0.26
D16Mit125	30	D16Mit47	42	70	39	0.25	0.09*
D19Mit68	3	D19Mit68	3	67	40	0.22	0.37*
DXMit114	41	DXMit5	67	67	32	0.32	0.27*

\*Intervals that deviate from Hardy-Weinberg equilibrium in both BXN and the advanced intercross.

intercross or backcross maps. When corrected using the Haldane-Waddington equation, the RI maps have a cumulative length of 1,400 cM, perfectly consistent with those of chromosome committee reports. Further improvements in the power and utility of RI strains will rely primarily on increased numbers and genetic diversity of these strains. Prospects are good, and more than 150 new mouse RI strains are currently being produced and genotyped by several research groups (see [32] for an updated list of investigators and new RI strains). For example, in collaboration with J.L. Peirce and L.M. Silver (Princeton University, USA), we are now producing over 40 new BXD RI strains. The first 20 lines have already been typed at over 600 markers. A set of approximately 85 RI strains has recently been completed by B. Bennett and T.E. Johnson (University of Colorado, Boulder, USA) and these lines are currently being genotyped at approximately 400 markers.

**Information content of RI strain sets**

Despite the accumulation of genotypes in RI strains, these genetic resources have often not been typed with sufficient density to accurately define the frequency and positions of recombination breakpoints. For example, in the venerable

set of 13 CXB strains, only 11 unique SDPs had been assigned to chromosome 1 before our work. With a more dense map of chromosome 1 that is now based on approximately 60 markers, we have recovered a total of 38 recombinations on this chromosome - approximately three recombinations per strain. The positions of these recombinations have been defined with a precision that ranges from 0.5 to 6.0 cM (2.3 cM average) as referenced to standard CCR maps. Twenty-one of the 38 SDPs are represented by one or more marker, but at least 17 SDPs remain to be defined and these SDPs unfortunately cannot be predicted unambiguously. For example, if two adjacent markers *P* and *D* have genotypes *BBCCC* and *CCCCC*, then there must be at least one unrecovered SDP between *P* and *D*. Until we actually type markers in the *P-D* interval, we do not know whether the intercalated SDP is *BCCCC* or *CBCCC*. To discover the undefined SDP could require considerable effort especially if available polymorphic markers on the *P-D* interval have been exhausted. All unrecovered SDPs lower the information content of an RI set. Their absence can significantly reduce linkage of both Mendelian and quantitative traits that are unlucky enough to be controlled by loci in the intervals with ambiguous SDPs.

How dense should a marker map be to define more than 90% of the total number of SDPs? With 862 markers, we were able to define approximately 60% of all likely SDPs among the 13 CXB strains. In the collected set of BXN RI strains, approximately 23% of the estimated 5,000 possible SDPs have been confidently defined with MIT microsatellites. We can estimate the density of the marker map that would be necessary to define 95% of all SDPs. For example, for the BXD set, if one assumes a random and independent distribution of breakpoints across strains and a random distribution of markers, it would take a map with about 2,700 markers to define 95% of the 1,536 SDPs.

**Use of the BXN set**

Most mapping software applications used by mouse geneticists are adapted for diallele crosses of various types. The BXN data set was therefore formatted in a way that collapses all non-B6 alleles into a single *N* class. The collected set of just over 100 strains can be used without complication with software such as Map Manager QTX [33,34]. This procedure was used largely as a convenience to integrate RI genetic maps. There are self-evident limitations that follow from the collapse of all non-*B* alleles (*A/J*, *DBA/2J*, *C3H/HeJ* and

**Figure 7** see figure on the next page

Correlation of genotypes illustrating non-syntenic associations for 102 strains. This sample from the complete correlation matrix of the BXN set illustrates both the expected syntenic correlations (the large red diagonal region extending down to the right) and several unexpected regions of high non-syntenic correlation between different chromosomes. Red regions are linked with positive correlation between 0.20 and 1.0 ( $p < 0.05$ ). Darker blue regions are linked with negative correlation of between -0.20 and -0.40 ( $p < 0.05$ ). Beige and light-blue regions are regions with intermediate correlation that are not statistically different from zero with 100 degrees of freedom. For example, the region of chromosome 1 near *D1Mit135* (labeled DIM 135 in this table) is linked positively to the proximal part of chromosome 19 and negatively to the proximal part of chromosome 2. The full data table is available online in several formats as Additional data files and at [30].



BALB/cByJ) into a single category. Geneticists using the BXN set should therefore begin virtually all studies by mapping with the individual component RI sets (AXB-BXA, BXD, BXH and CXB) to detect possible levels of allele effects (an allelic series). The *B* allele is a common feature and may be a useful reference point for estimating hierarchies among the five parental alleles. This separate, set-by-set analysis prevents the *N* alleles from averaging out, as they might in a cumulative analysis (the *N* alleles will often have effects that are both higher and lower than that of the *B* allele). Because the BXN set includes 490 common marker loci and a consistent alignment and integration of the component RI maps, it is now much easier to combine linkage likelihood ratios from the component RI sets. A simple method based on Fisher's method is described by Williams and colleagues [8] in a study that pooled data from BXD and BXH sets. More sophisticated methods for automatically extracting and combining linkage statistics from the multi-allele BXN sets will require modification of mapping application programs. Pooling data in this way will require judicious and well justified statistical procedures. Combining data across the BXN sets can easily degrade a linkage analysis. The statistical exploration of different combinations of RI sets provides new degrees of freedom which may generate false-positive results, but which may also generate interesting hypotheses regarding QTL action.

The BXN map could be refined further by interpolating genotypes of other markers and genes that have been mapped independently by many investigators in single RI sets. Our BXD database includes only microsatellite loci, for example, and excludes hundreds of potentially informative polymorphic loci, many in interesting genes. We regret having to use this procrustean approach, but because of the difficulty of verifying genotypes and because numerous loci introduce improbable double-recombinant haplotypes, we have used exclusive criteria to ensure high-quality maps. Investigators interested in recovering some of this lost data should refer to the comprehensive lists of genotypes maintained by the Mouse Genome Database [35]. However, genotypes of any marker and strain that introduce new double-recombinants into the BXN map should be regarded with a high level of suspicion.

#### **Power and precision of RI strains**

A set of 100 conventional RI strains will have twice the genetic variance of a matched set of 100 F2 progeny and four times that of 100 backcross progeny. This increased genetic variance comes at some cost: 100 F2 animals represent 200 meioses and contain almost 200 unique haplotypes per chromosome (the non-recombinant chromosomes reduce this number somewhat). RI strains are fully inbred and 100 lines represent almost 100 unique haplotypes per chromosome. A set of 100 RI strains therefore has approximately twice the load of recombinations as 100 F2s. For a semidominant Mendelian trait, 100 RI strains therefore provide

roughly twice the precision of 100 F2 progeny and four times that of 100 N2 progeny. When both genetic variance and recombination load are considered together, a set of 100 RI strains should be approximately four times as effective (precise) for mapping complex traits as an F2, and eight times as effective as a backcross. This estimate assumes that only a single RI animal is sampled per line; a strategy that is appropriate for mapping SNPs, microsatellites and other Mendelian loci.

The gain for mapping quantitative traits will be greater and will depend strongly on the heritability and to a lesser extent on the degree of dominance at each locus. Belknap [3] has compared the relative power of RI strains and F2 intercrosses under several models and assuming different levels of heritability. For morphometric traits such as brain weight, with narrow sense heritabilities of around 0.5, 100 RI strains will provide a level of precision and power that is conservatively equivalent to that of 600-1,000 F2 intercross progeny. The advantage shifts further in favor of RI strains for traits with lower heritability. Power is one key issue in QTL mapping, but at present, precision - the ability to fine-map QTLs to subcentimorgan intervals suitable for candidate gene analysis - is the hurdle, and one that would be less imposing with improved RI resources [36].

#### **Making better RI resources**

The usefulness of RI strains for mapping is largely a function of the number of known recombination breakpoints and useful polymorphisms that they harbor. All current mouse RI sets are small, and consequently the most common criticisms leveled at QTL mapping with RI strains is that the precision and power are poor and that only those QTLs with unusually large effects can be detected reliably. The BXN set provides only a partial solution to this problem by expanding the set of RI strains that can be treated statistically as a complex cross. A much better long-term solution is to generate larger sets of RI strains for high-precision complex trait analysis. RI sets consisting of 100 to 1,000 lines could provide very impressive power and subcentimorgan precision. The LXS set (80-90 strains) and the enlarged BXD set (70-80 strains) mentioned above will soon provide practical demonstrations. Generating large sets is an undertaking, but the effort is dwarfed by ongoing mutagenesis and sequencing efforts. Generating, maintaining and storing 1,000 RI lines could be a well justified expense given the long-term utility of large RI sets in tackling otherwise intractable problems in functional genomics - gene pleiotropy, genetic correlations, epistasis and reaction-norm genetics - in a mammal.

Several other factors make this idea significantly more attractive. First, an RI set can be produced using more than two inbred strains. Four to eight strains could in principle be combined to make RI sets that segregate for a greater variety of polymorphisms. This addresses the concern that a single conventional diallele RI set may not be useful for studying

particular traits because of a paucity of relevant polymorphisms. Such multi-way RI lines buck the reductionist trend of eliminating genetic complexity by isolating gene variants on inbred backgrounds, but such complexity has its advantages and these strains would provide welcome models for exploring genetic background effects that plague much of the current work on transgenic and knockout mice [37]. Second, by genotyping and selectively breeding the most highly recombinant animals it should be possible to generate RI strain sets with map expansions that significantly exceed that predicted by the Haldane-Waddington equation, an equation that assumes random mating of sibs. A six- to eight-fold expansion should be attainable, particularly if recombinations are tracked before and during the inbreeding process (Figure 4).

Recombination density could be further increased by starting RI strains from either advanced intercross progeny [36] or heterogeneous stock (Figure 4) as was done in making the new set of 40 BXD strains mentioned above. Third, the power of RI sets can now be amplified significantly by use of RI intercross (RIX) and RI backcross (RIB) designs [19,20]. Finally, large RI sets will largely eliminate the problem of non-syntenic association.

A second well justified objection to using RI strains to map quantitative traits is that fully inbred strains may not provide representative phenotypes precisely because they are inbred and subject to often severe inbreeding depression. The abnormal genetic architecture of inbred strains and the fixation of multiple alleles that affect fitness will almost inevitably produce unusual pleiotropic and epistatic effects on a range of complex traits. Outliers are common on these and other inbred lines. Can the strain means be trusted?

RIX progeny provide a surprisingly simple solution to this problem [19,20]. RIX progeny made among members of a single diallele RI set will be similar to an F2 intercross with an inbreeding coefficient of 0.5. Crosses between members of completely different RI sets (for example, AXB1 crossed to LXS80) will have an inbreeding coefficient close to zero. In this respect they will be more appropriate models of human genetic variation, but with the remarkable advantages of completely defined genotypes and the option of generating large numbers of isogenic individuals.

### Using the BXN and their RIX progeny

QTLs mapped using RI sets can be quickly verified and positionally refined by generating sets of RIX and RIB lines between those parental strains that have recombinations in critical QTL intervals. The RIX method has already proved a highly effective way of extracting QTLs from the tiny set of 13 CXB strains [19,20]. The 13 inbred lines have the potential to be converted to as many as 156 F1 lines, of which small subsets can be selected based on parental genotypes to test particular candidate QTLs and to simultaneously recover

gene dominance signal by generating F1 heterozygotes. This greatly increases the power to detect QTLs in the presence of strong genetic, parental and developmental background noise, and at the same time exposes gene dominance deviations to help refine QTL effect and position. The BXN opens up a huge RIX domain for analysis. Approximately 88 BXN RI strains are now available from the Jackson Laboratory, and these strains can be crossed to generate about  $88 \times 87/2$  (3,828) genetically unique recombinant inbred intercross progeny (RIX progeny) with breakpoints in precisely defined intervals. Each one of these F1s can be made in reciprocal pairs to assess the role of parental effects (for example, a BXD1 mother crossed to an AXB2 father or vice versa) and, like RI strains, many isogenic individuals can be typed to reduce non-genetic variance.

Selected subsets of this huge pool of 3,828 unique RIX genomes can be made by crossing those RI strains with breakpoints in intervals thought to harbor QTLs. These interval-specific RIX progeny can be phenotyped and used to refine the genetic analysis of complex traits. Once QTLs have been mapped to candidate intervals, the subset of strains with recombinations within those intervals becomes an important resource for confirming and refining QTL location [33]. This is especially the case if one exploits the RIX method. For example, if a QTL maps between 10 and 25 cM on chromosome 1 in the BXD set (that is between *D1Mit430* and *D1Mit375*), and if *B* alleles in this interval are associated with high phenotypes, then the cross of BXD15 with BXD20 may be particularly informative because the F1 hybrid is an obligatory *B* homozygote on a short interval between 15 cM and 17 cM and is also an obligatory *D* homozygote proximal to 13 cM and distal to 18 cM. A set of isogenic F1 RIX progeny made by crossing several RI lines with recombinations in a critical interval can be used to refine the probable position of a QTL. Map Manager QTX has now been updated to automatically generate the genotypes of the RIX progeny produced by a one-generation cross of RI parents [34]. Given this huge sample of unique RIX genomes, even modest quantitative differences between C57BL/6 and other strains should be readily mapped (or confirmed) using the BXN and RIX mapping.

### Mapping modifiers of dominant alleles using RI backcrosses

Knowing the precise location of breakpoints in RI lines also makes it possible to map modifier loci of mutations by making and phenotyping a set of different F1 crosses made between inbred carrier stock (for example, a knockout carried on a C57BL/6 background) and fully typed RI lines. A set of these RI backcrosses (RIB) has a genetic structure similar to a conventional N2 backcross, but there is no need to genotype any of the RIB progeny and they have the major advantage that isogenic progeny can be typed to obtain much more reliable trait scores. This method does depend on either a dominant or semidominant mutant allele, since the phenotype must be detectable on a significant fraction of

the RIB progeny. Provided that this condition is met, the costs and logistics of this type of screen may be more modest than a typical screen for modifier loci. The analysis can be carried out without genotyping and using replicated genomes to test for environmental modulators.

### BXN and sequencing efforts

Five of the widely used sets of RI strains that we have typed and analyzed share C57BL/6 as a parental strain. The genome of C57BL/6J is currently being sequenced as part of a public effort [38] and for this reason, the utility of the BXN set for converting QTLs to strong candidate genes will increase significantly in the next few years [37]. It will become far easier to generate complete lists of positional candidate genes and then to obtain data on gene and protein expression patterns. The two other major strains incorporated into the BXN set - A/J and DBA/2J - are also being sequenced by Celera Genomics and, in principle, it will be possible to compare sequences of these three major strains to generate lists of possible allelic variants in positional candidate genes. The recent cloning of the *Sac* locus that controls sugar and saccharin preference on distal chromosome 4, provides a good example of the increased power of candidate gene analysis. This locus was initially mapped using 20 BXD strains [39,40]. In the absence of high-resolution mapping, but with astute analysis of human and mouse sequence data, *Sac* was identified almost simultaneously by several groups as the gene for the T1R3 receptor [41-47]. In a few years, the identification of genes associated with QTLs will probably be no more of a special exception than the cloning of Mendelian genes was in the mid-1990s.

## Materials and methods

### Strains and DNA

Genomic DNA from most recombinant inbred and parental strains was purchased from the Jackson Laboratory, Bar Harbor, USA. DNA was obtained from 40 of 41 AXB and BXA strains and 35 of 36 BXD strains, 13 CXB strains, and 12 BXH strains - 100 strains total. For visual clarity in this paper we have dropped hyphens and substrain designations from RI strain names. For example, strain BXD-1/Ty is referred to as BXD1. Databases and web-accessible data tables at the Informatics Center for Mouse Neurogenetics [21] also use this simplified nomenclature.

All DNA from the Jackson Laboratory Mouse DNA Resource was extracted from individual male mice. The RI animals that we genotyped were, with a few exceptions, the progeny of more than 20 serial matings between siblings. Data on the particular generation that we used for genotyping and the current generation of RI animals is available at [30]. DNA from seven new BXH strains generated by Linda Siracusa (Thomas Jefferson Medical College, Philadelphia, USA) was extracted from the spleen using a high-salt procedure [48]. The new BXH strains were generated by crossing C57BL/6J-*c2J/c2J* albino males with

C3H/HeJ females and their production and genotyping will be described in detail elsewhere (L. Siracusa and R.W.W., unpublished data). Three of the new BXH albino strains are no longer available (C2, D1 and E2). We genotyped 107 RI strains. Several sets of strains share haplotypes (Table 8). We deleted redundant strains (AXB18, ABX20 and BXA17).

Strains BXHD1, BXHE1 and BXHE2 were backcrossed to C57BL/6J for one generation before sib-matings were begun. There is therefore a pronounced increase in the number of chromosomal segments inherited from C57BL/6J. These N2-derived RI strains were dropped from most aspects of the analysis of RI genome structure. BXD41 has been extinct for several years and was never completely inbred. Although we have DNA for this strain, our sample is from a F12 generation male. We did not genotype BXD41 in this study.

We refer to the collected RI set as the BXN set because each of the strains includes C57BL/6 (B6 or *B*) as one of the parental strains - the common substrain C57BL/6J in the case of AXB, BXA, BXD and BXH, and the substrain C57BL/6By in the case of CXB. The other parental strain in the BXN set is not B6-derived: A/J in both AXB and BXA sets, DBA/2J in BXD, C3H/HeJ in BXH and BALB/cBy in CXB.

### PCR

Microsatellite loci distributed across all autosomes and the X chromosome were typed using a modified version of the protocol of Love and colleagues [49] and Dietrich and colleagues [23] described in detail at [50]. A total of 1,773 primer pairs (MapPairs) that selectively amplify polymorphic MIT microsatellite loci were purchased from Research Genetics. Each 10  $\mu$ l PCR reaction mixture contained 1x PCR buffer, 1.92 mM MgCl<sub>2</sub>, 0.25 units Taq DNA polymerase, 0.2 mM of each deoxynucleotide, 132 nM of the primers, and

**Table 8**

#### The strains that have been genotyped in this study

RI name	Strain name	Total N	Available N
AXB	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 17, (18 = 19 = 20), 21, 23, 24	19	12
BXA	1, 2, 4, 7, (8 = 17), 9, 11, 12, 13, 14, 16, 18, 20, 22, 23, 24, 25, 26	17	13
BXD	1, 2, 5, 6, 8, 9, 11, 12 to 16, 18 to 25, 27 to 36, 37, 38 to 40, 41, 42	36	34
BXH	2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 14, 19, A1, A2, B2, C2, D1, E1, E2	19	16
CXB	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	13

Several sets of strains in parentheses share haplotypes. Strains printed in italic are extinct or are available only from cryopreserved stock.



50 ng genomic DNA. Reactions were set up using a 96-channel pipetting station. A loading dye (60% sucrose, 1.0 mM cresol red) was added to the reaction before the PCR [51]. PCRs were carried out in 96-well microtiter plates. We used a high-stringency touchdown protocol in which the annealing temperature was lowered progressively from 60°C to 50°C in 2°C steps over the first six cycles [52]. After 30 cycles, PCR products were run on cooled 2.5% Metaphor agarose gels (FMC Inc., Rockland ME), stained with ethidium bromide, and photographed. Gel photographs were scored and directly entered into relational database files.

Eighteen primer pairs were resynthesized at our request by Research Genetics using the original sequence data (Whitehead/MIT SSLP Database Release 8 [53]) to verify that our chromosome reassignments of microsatellite loci were not due to the use of incorrect primer sequences.

### Common markers

When we began this work fewer than 25 MIT markers had been typed on each of the four major RI sets. We were able to increase this to 489 markers. We relied on these loci to assemble consensus RI maps. The additional 986 MIT markers were typed by us and other groups in at least one set of RI strains. The BXN genotype database includes 1,578 markers. Any pair of RI sets share between 500 and 600 fully genotyped markers. For example, the two largest RI sets - AXB-BXA and BXD - have been typed at 591 common microsatellite markers.

### Databases

Relational database files were assembled from the 1998-2000 chromosome committee reports, the Portable Dictionary of the Mouse Genome [54] and the MIT/Whitehead SSLP database Release 8 [53]. These files contain a summary of information on chromosomal positions of 6,332 MIT microsatellite markers and information on an additional 15,000 genes and markers. We have included Nuffield Department of Surgery (Nds) microsatellite markers for which primer sequences are available. Additional databases devoted to each RI set were assembled from text files downloaded from the Mouse Genome Database [35]. New and corrected genotypes were entered directly into these files.

### Additional data files

Additional data files available online include versions of the BXN genetic maps and microsatellite marker genotypes and the two-locus correlation matrices of genotypes for different subsets of strains [30].

### Acknowledgements

This project was supported by a Human Brain Project/Neuroinformatics program project (Informatics Center for Mouse Neurogenetics) funded jointly by the National Institute of Mental Health, National Institute on Drug Abuse, and the National Science Foundation (P20-MH 62009). We

thank Xiyun Peng for assistance in genotyping CXB and BXH mice. We also thank Felisha Scruggs at Research Genetics Inc. (Invitrogen) for resynthesizing 18 primer pairs for our use, Susan Deveau of the Jackson Laboratory DNA Resource for information on the generation numbers of RI DNA samples, and David Threadgill, Gary Churchill and Kenneth Manly for comments on drafts of this paper.

### References

1. Taylor BA: **Recombinant inbred strains**. In *Genetic Variants and Strains of the Laboratory Mouse*, 2nd edn. Edited by Lyon ML, Searle AG. Oxford: Oxford University Press; 1989: 773-796.
2. Bailey DW: **Strategic uses of recombinant inbred and congenic strains in behavior genetics research**. In *Genetic Research Strategies for Psychobiology and Psychiatry*. Edited by Gershon ES, Matthyse S, Breakefield XO, Ciaranello ED. New York: Plenum; 1981: 189-198.
3. Belknap JK: **Effect of within-strain sample size on QTL detection and mapping using recombinant inbred strains of mice**. *Behav Genet* 1998, **28**:29-38.
4. Crabbe JC, Wahlsten D, Dudek BC: **Genetics of mouse behavior: interactions with laboratory environment**. *Science* 1999, **284**:1670-1672.
5. Toth LA, Williams RW: **A quantitative genetic analysis of slow-wave sleep in influenza-infected CXB recombinant inbred mice**. *Behav Genet* 1999, **29**:339-348.
6. Hain HS, Crabbe JC, Bergeson SE, Belknap JK: **Cocaine-induced seizure thresholds: quantitative trait loci detection and mapping in two populations derived from the C57BL/6 and DBA/2 mouse strains**. *J Pharmacol Exp Ther* 2000, **293**:180-187.
7. Williams RW: **Neuroscience meets quantitative genetics: using morphometric data to map genes that modulate CNS architecture**. In *Short Course in Quantitative Neuroanatomy*. Edited by Morrison J, Hof P. Washington DC: Society of Neuroscience; 1998: 66-78.
8. Williams RW, Strom RC, Goldowitz D: **Natural variation in neuron number in mice is linked to a major quantitative trait locus on Chr 11**. *J Neurosci* 1998, **18**:138-146.
9. Strom RC, Williams RW: **Cell production and cell death in the generation of variation in neuron number**. *J Neurosci* 1998, **18**:9948-9953.
10. Williams RW: **Mapping genes that modulate mouse brain development: a quantitative genetic approach**. In *Mouse Brain Development*. Edited by Goffinet A, Rakic P. Berlin: Springer; 2000: 21-49.
11. Williams RW, Airey DC, Kulkarni A, Zhou G, Lu L: **Genetic dissection of the olfactory bulb of mice: QTLs on chromosomes 4, 6, 11, and 17 modulate bulb size**. *Behav Genet* 2001, **31**:61-77.
12. Lu L, Airey DC, Williams RW: **Complex trait analysis of the hippocampus: Mapping and biometric analysis of two novel gene loci with specific effects on hippocampal structure in mice**. *J Neurosci* 2001, **21**:3503-3514.
13. Airey DC, Lu L, Williams RW: **Genetic control of the mouse cerebellum: localization of quantitative trait loci modulating size and architecture**. *J Neurosci* 2001, **21**:5099-5109.
14. Rosen GD, Williams RW: **Complex trait analysis of the mouse striatum: independent QTLs modulate volume and neuron number**. *BMC Neuroscience* 2001, **2**:5. [<http://www.biomedcentral.com/1471-2202/2/5>]
15. Weber JL, Broman KW: **Genotyping for human whole-genome scans: past, present, and future**. *Adv Genet* 2001, **42**:77-96.
16. **Mouse Genome Informatics Release 2.5** [[www.informatics.jax.org/searches/riset\\_form.shtml](http://www.informatics.jax.org/searches/riset_form.shtml)]
17. Sampson SB, Higgins DC, Elliot RW, Taylor BA, Lueders KK, Koza RA, Paigen B: **An edited linkage map for the AXB and BXA recombinant inbred mouse strains**. *Mamm Genome* 1998, **9**:688-694.
18. Taylor BA, Wnek C, Kotlus BS, Roemer N, MacTaggart T, Phillips SJ: **Genotyping new BXD recombinant inbred mouse strains and comparison of BXD and consensus maps**. *Mamm Genome* 1999, **10**:335-348.
19. Threadgill DW, Airey DC, Lu L, Manly K, Williams RW: **Recombinant inbred intercross mapping: a new approach to extend the power of existing mouse resources**. *Int Mouse Genome Conf* 2001, **15**: in press.

20. Williams RW, Threadgill DW, Airey DC, Gu J, Lu L: **RIX Mapping: a demonstration using CXB RIX hybrids to map QTLs modulating brain weight in mice.** *Soc Neurosci Abst* 2001, **27**: in press.
21. **Informatics Center for Mouse Neurogenetics: Mouse Genome Databases** [www.nervenet.org/main/dictionary.html]
22. Jeffreys AJ, Wilson V, Kelly R, Taylor BA, Bulfield G: **Mouse DNA "fingerprints": analysis of chromosome localization and germ-line stability of hypervariable loci in recombinant inbred strains.** *Nucleic Acids Res* 1987, **15**:2823-2836.
23. Dietrich WF, Katz H, Lincoln SE: **A genetic map of the mouse suitable for typing in intraspecific crosses.** *Genetics* 1992, **131**:423-447.
24. Nelson JC: **QGENE: software for maker-based genomics analysis and breeding.** *Mol Breeding* 1997, **3**:239-245.
25. **QGene** [http://www.qgene.com]
26. Haldane JBS, Waddington CH: **Inbreeding and linkage.** *Genetics* 1931, **16**:357-374.
27. Lynch M, Walsh B: *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates; 1998.
28. **Mammalian Genotyping Service** [http://research.marshfield-clinic.org/genetics/Genotyping\_Service/mgsver2.htm]
29. **Mouse genetic map files** [http://www.nervenet.org/MMfiles/MMlist.html]
30. **Informatics Center for Mouse Neurogenetics: publications** [www.nervenet.org/papers/bxn.html]
31. Eleftheriou BE, Elias MF, Castellano C, Oliverio A: **Cortex weight: a genetic analysis in the mouse.** *J Hered* 1975, **66**:207-212.
32. **A Clearinghouse for Information on Recombinant Inbred Strains** [http://brain.utm.edu/clear/phps/menu.php]
33. Manly KF, Olson JM: **Overview of QTL mapping software and introduction to Map Manager QT.** *Mamm Genome* 1999, **10**:327-334.
34. **Map Manager QTX** [http://mapmgr.roswellpark.org/mmQTX.html]
35. **Mouse Genome Database** [http://www.informatics.jax.org/]
36. Darvasi A: **Experimental strategies for the genetic dissection of complex traits in animals.** *Nat Genet* 1998, **18**:19-24.
37. Lariviere WR, Chesler EJ, Mogil JS: **Transgenic studies of pain and analgesia: mutation or background genotype?** *J Pharm Exp Ther* 2001, **297**:467-473.
38. Marshall E: **Celera assembles mouse genome; public labs plan new strategy.** *Science* 2001, **292**:822-823.
39. Belknap JK, Crabbe JC, Phillips TJ, Hitzemann R, Buck KJ, Williams RW: **Quantitative trait loci and genome-wide mutagenesis: two phenotype-driven approaches to the dissection of complex murine traits.** *Behav Genet* 2001, **31**:5-15.
40. Belknap JK, Crabbe JC, Plomin R, McClearn GE, Sampson KE, O'Toole LA, Gora-Maslak G: **Single-locus control of saccharin intake in BXD/Ty recombinant inbred (RI) mice: some methodological implications for RI strain analysis.** *Behav Genet* 1992, **22**:81-100.
41. Phillips TJ, Crabbe JC, Metten P, Belknap JK: **Localization of genes affecting alcohol drinking in mice.** *Alcohol Clin Exp Res* 1994, **18**:931-941.
42. Li X, Inoue M, Reed DR, Huque T, Puchalski RB, Tordoff MG, Ninomiya Y, Beauchamp GK, Bachmanov AA: **High-resolution genetic mapping of the saccharin preference locus (Sac) and the putative sweet taste receptor (TIR1) gene (Gpr70) to mouse distal chromosome 4.** *Mamm Genome* 2001, **12**:13-16.
43. Montmayeur JP, Liberles SD, Matsunami H, Buck LB: **A candidate taste receptor gene near a sweet taste locus.** *Nat Neurosci* 2001, **4**:492-498.
44. Max M, Shanker YG, Huang L, Rong M, Liu Z, Campagne F, Weinstein H, Damak S, Margolskee RF: **Tas1r3, encoding a new candidate taste receptor, is allelic to the sweet responsiveness locus Sac.** *Nat Genet* 2001, **28**:58-63.
45. Sainz E, Korley JN, Battey JF, Sullivan SL: **Identification of a novel member of the TIR family of putative taste receptors.** *J Neurochem* 2001, **77**:896-903.
46. Kitagawa M, Kusakabe Y, Miura H, Ninomiya Y, Hino A: **Molecular genetic identification of a candidate receptor gene for sweet taste.** *Biochem Biophys Res Commun* 2001, **283**:236-242.
47. Bachmanov AA, Li X, Reed DR, Ohmen JD, Li S, Chen Z, Tordoff MG, de Jong P, Wu C, West DB, *et al.*: **Positional cloning of the mouse saccharin preferences (Sac) locus.** *Chem Senses* 2001, **26**:925-933.
48. Laird PW, Zijderfeld A, Linders K, Rudnicki M, Jaenisch R, Berns A: **Simplified mammalian DNA isolation procedure.** *Nucleic Acids Res* 1991, **19**:4293.
49. Love JM, Knight AM, McAleer MA, Todd JA: **Towards construction of a high resolution map of the mouse genome using PCR-analyzed microsatellites.** *Nucleic Acids Res* 1990, **18**:4123-4130.
50. **Informatics Center for Mouse Neurogenetics: Publications** [http://www.nervenet.org/papers/pcr.html]
51. Routman E, Cheverud J: **A rapid method of scoring simple sequence repeat polymorphisms with agarose gel electrophoresis.** *Mamm Genome* 1994, **5**:187-188.
52. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS: **"Touch-down" PCR to circumvent spurious priming during gene amplification.** *Nucleic Acids Res* 1991, **19**:4008.
53. **Whitehead/MIT SSLP Database Release 8** [http://www-genome.wi.mit.edu/genome\_data/mouse/mapbase\_query\_form.html]
54. Williams RW: **The Portable Dictionary of the Mouse Genome: a personal database for gene mapping and molecular biology.** *Mamm Genome* 1994, **5**:372-375.
55. Panoutsakopoulou V, Spring P, Cort L, Sylvester JE, Blank KJ, Blankenhorn EP: **Microsatellite typing of CXB recombinant inbred and parental mouse strains.** *Mamm Genome* 1997, **8**:357-361.