Interspecific recombinant congenic strains between C57BL/6 and mice of the *Mus spretus* species : a powerful tool to dissect genetic control of complex traits.

Gaëtan Burgio[*,†], Marek Szatanik[‡], Jean-Louis Guénet[*], Maria-Rosa Arnau[$], Jean-Jacques Panthier[*], and Xavier Montagutelli[*,1]

[*] Unité postulante de Génétique fonctionnelle de la Souris, CNRS URA 2578, Institut Pasteur, 75015 Paris, France

[†] Unité d'éco-anthropologie, équipe « génétique des populations humaines », CNRS MNHN, P7, UMR 5145, Musée de l'Homme, 75016 PARIS, France.

[‡] Unité des Neisseria, Institut Pasteur, 75015 Paris, France

[$] Universidad de La Laguna, 38207 Teneriffe, Spain

Running title : QTL analysis in mouse interspecific RCS

Key-words : Mouse, QTL, Interspecific Recombinant Congenic Strains, Complex traits, *Mus spretus*

[1] Corresponding author :

Xavier Montagutelli

Unité postulante de Génétique fonctionnelle de la Souris

25 rue du Docteur Roux, 75724 PARIS cedex 15, France

Tel: (33) 1 45 68 89 55 - Fax: (33) 1 45 68 86 34

email: xmonta@pasteur.fr

ABSTRACT

Complex traits are under the genetic control of multiple genes, often with weak effects and strong epistatic interactions. We developed two new collections of mouse strains to improve genetic dissection of complex traits. They are derived from several backcrosses of the *Mus spretus* SEG/Pas or STF/Pas strains on the C57BL/6J background. Each of the 55 Interspecific Recombinant Congenic Strains (IRCS) carries up to eight SEG/Pas chromosomal segments with an average size of 11.7 Mb, totalizing 1.37% of the genome. The complete series covers 39.7% of the SEG/Pas genome. As a complementary resource, six partial or complete interspecific consomic strains were developed, and increase genome coverage to 45.6%. To evaluate the usefulness of these strains for QTL mapping, 16 IRCS were compared with C57BL/6J for seven hematological parameters. Strain 66H, which carries three SEG/Pas chromosomal segments, had lower red blood cell volume and higher platelet count than C57BL/6J. Each chromosomal segment was isolated in a congenic strain to evaluate individual effects. Congenic strains were combined to assess epistasis. Our data show that both traits were controlled by several genes with complex epistatic interactions. IRCS are therefore useful to unravel QTLs with small effects, and gene by gene interactions.

INTRODUCTION

Individuals of every species exhibit more or less profound differences which affect all aspects of their anatomy, physiology, development, behavior, and susceptibility to various diseases. Most of these phenotypic differences are under the control of multiple genes and environmental factors, and are therefore referred to as complex traits, by contrast with monogenic traits, where a single gene modification controls phenotypic variation. Complex traits are assessed in individuals as one or more quantitative measurements, so that genes controlling these variations are quantitative trait loci, or QTLs. Several noticeable features of QTLs which make their identification difficult are the lack of strict relationship between phenotypes and genotypes, low penetrance, epistasis and pleiotropy.

Because of their genetic architecture, complex traits require that sophisticated mapping populations be developed and studied (ABIOLA *et al.* 2003; FLINT *et al.* 2005). Model organisms such as the mouse offer the advantage of providing conditions resembling human pathology (e.g. diabetes mellitus, asthma, hypertension, and atherosclerosis) together with the ability to produce segregating populations optimized for the dissection and functional analysis of the underlying genes.

Complex traits are often analyzed using two-generation crosses, such as backcrosses and F2s. While such crosses can be produced between almost any pair of inbred strains, their analysis results in low power to detect weak QTLs, produces large confidence intervals (DARVASI *et al.* 1993), and require that presumptive QTLs be confirmed in congenic strains (ABIOLA *et al.* 2003). Heterogeneous stocks (HS) represent an outbred population produced from an eight-way intercross followed by several generations of random breeding to increase the density of recombination events (MOTT and FLINT 2002). This strategy has yielded the identification of an incredibly large number of QTLs with very high LOD scores and

extremely small confidence intervals (VALDAR *et al.* 2006). However, neither two-generation crosses nor HS can be perpetuated, hence preventing further studies, on the same population.

Recombinant inbred strains are permanent resources derived from an F2 cross (BAILEY 1971; BAILEY 1981; TAYLOR 1978). Their inbred status allows replication of phenotype assessment in genetically identical individuals, and for correlations between traits which cannot be measured in the same individuals. However, most existing sets, which were developed between a restricted number of laboratory inbred strains, are too small to provide the required power for QTL detection and fine mapping, especially when small-effect QTLs are involved, with some degree of epistasis (FLINT *et al.* 2005). To overcome these limitations, a much larger reference population is currently under development. The Collaborative Cross aims of producing a series of over 1,000 inbred strains with a balanced contribution from eight parental inbred strains, two of which are derived from a different subspecies (CHURCHILL *et al.* 2004). Consomic strains are an ordered series of inbred strains which share the same inbred genetic background and differ by only one entire chromosome which has been introgressed from another inbred strains by repeated backcrosses (SINGER *et al.* 2004). While any phenotypic variation between a consomic strain and the background strain can readily be attributed to one or more genes on the differential chromosome, epistatic interaction between non-syntenic genes may be missed. Other genetic tools developed for genetic analysis of complex traits include sets of overlapping congenic strains covering the entire genome (DAVIS *et al.* 2005; IAKOUBOVA *et al.* 2001).

Recombinant congenic strains (RCS) were introduced by Demant and colleagues (DEMANT and HART 1986) as inbred populations derived after two or three backcrosses, resulting in unequal genomic contribution from the two parental inbred strains. With three backcrosses, each RCS is expected to carry 12.5% (1/8) of its genome from the "donor" strain. Theoretical considerations predict that about 95% of the genome of the donor strain

can be transferred and distributed within a series of 20 RCS (DEMANT and HART 1986; MOEN *et al.* 1997). Five sets of such RCS have been produced between pairs of laboratory inbred strains (FORTIN *et al.* 2001b; GROOT *et al.* 1992; MARTIN *et al.* 1992b) and the results have fulfilled the theoretical expectations (STASSEN *et al.* 1996). Since each RCS carries only a small fraction of the donor genome, multiple genes controlling a trait are likely to be isolated in different strains (DEMANT and HART 1986; MOEN *et al.* 1991). QTL mapping is achieved in an F2 population produced between the "background" strain and an RCS with a contrasting phenotype (VAN WEZEL *et al.* 1996). Because only one eighth of the genome is segregating in such a cross, the total genetic variance is reduced and the power to identify QTLs is higher, as is the ability to detect epistatic interactions since a lower value of the statistics is required to declare significant an interaction between two loci. Recombinant congenic strains have been successfully used to study a variety of complex traits (BANUS *et al.* 2005; FORTIN *et al.* 2001a; GOUYA *et al.* 2007; LEE *et al.* 2006; LEMAY and HASTON 2005; MARTIN *et al.* 1992a; MOEN *et al.* 1996; MOEN *et al.* 1992; ROY *et al.* 2006; SERREZE *et al.* 1994).

Besides genetic populations of appropriate structure, genetic analysis of complex trait also requires phenotypic polymorphism. While most genetic tools have been developed using laboratory mice, wild-derived inbred strains have proven highly valuable in providing original genotypic and phenotypic variations (GUENET and BONHOMME 2003; GUÉNET and MONTAGUTELLI 1994). They offer a much higher level of diversity (IDERAABDULLAH *et al.* 2004), and polymorphism rate at the sequence level is estimated to be 1-1.5% between C57BL/6J and *Mus spretus*-derived strains, such as SEG/Pas (GUENET and BONHOMME 2003), the same order of magnitude as that between human and chimpanzee (NEWMAN *et al.* 2005). Interspecific and intersubspecific crosses have been at the origin of high-density, high-resolution genetic maps (AVNER *et al.* 1988; RHODES *et al.* 1998). Wild-derived inbred strains also differ from laboratory mice in their susceptibility to various traits (MELANITOU *et al.*

1998; STAELENS *et al.* 2002; YI *et al.* 2004), including viral (MASHIMO *et al.* 2002), bacterial (SEBASTIANI *et al.* 2002; TURCOTTE *et al.* 2006), and parasitic (BAGOT *et al.* 2002) infections. It has been suggested however that, because of this high genetic polymorphism, phenotypic differences might be under the control of many QTLs which could hamper the identification of the genes.

In an attempt to combine the power of interspecific crosses and that of recombinant congenic strains, we developed a set of interspecific recombinant congenic strains (IRCS), designated BcG, with the original aim of introgressing 12.5% of *Mus spretus* genome (strain SEG/Pas) in a C57BL/6J inbred background. A total of 55 strains were derived from three to five backcrosses, bred to homozygosity, and genotyped for 673 genetic markers spread over the genome. Surprisingly, the final proportion of *Mus spretus* genome retained in each strain varied from 0 to 3.8%, with an average of 1.37%. This six-fold reduction compared to expectations based on the actual crosses was due to strong selection against *Mus spretus* alleles in the course of inbreeding generations. In parallel, we attempted to develop a set of interspecific consomic strains (ICS) from the same strain combination. Only fractions of six chromosomes could be retained. Altogether, IRCS and ICS cover 45.6% of the genome. We have analyzed a few IRCS for several quantitative traits and we demonstrate that they offer a useful tool to explore the complexity of the genetic control of multigenic traits.

MATERIALS AND METHODS

Mice and crosses

IRCS were developed at the Institut Pasteur in Paris. C57BL/6JIco (B6) mice were purchased from IFFA-CREDO (now Charles River Laboratories France, L'Arbresle, France). The SEG/Pas colony was established from breeders provided by François Bonhomme (Montpellier, France) and maintained by relaxed inbreeding (closed colony).

SEG/Pas males were mated to B6 females to produce F1 females. F1 females were mated to B6 males to produce a N2 generation which became part of the EUCIB mapping panel (BREEN *et al.* 1994; RHODES *et al.* 1998). N2 males (N=157) were mated to two OF1 outbred females IFFA-CREDO (L'Arbresle, France) to assess their fertility. Eighteen fertile males were then mated to B6 females to yield N3 progeny. Between four and eight N3 breeding pairs were established from the progeny of each N2 male, as the starting point for recombinant congenic strains (Figure 1). As many as four breeding pairs or trios were set with the progeny of each N3 intercross. In the following generations, each strain was maintained as a set of four breeding pairs or trios. Whenever possible, mating involved full brothers and sisters.

Approximately half of the strains derived from N3 mice stopped breeding after one to four generations of inbreeding. About half of them were definitely lost. For the other half, one additional backcross to B6 was performed by mating one or more males of the strain to B6 females. Inbreeding was restarted from their progeny.

We initially estimated the number of IRCS needed to cover the genome as follows. Let K be the number of N2 males used as progenitors, N the total number of strains, N_i the number of strains derived from N2 male i ($N = \Sigma\ Ni$), and P the probability that all N IRCS are homozygous for the B6 allele at a given locus. The probability that a given strain will be B/B at a locus is the sum of two probabilities: the probability that the N2 male was B/B and the

probability that it was B/S but the *Mus spretus* allele was not fixed after inbreeding. Because N2 males were crossed again to B6 females before inbreeding starts, $P = 0.5 + 0.5 \times (0.5 + 0.5 \times 0.5) = 0.875$. If all strains were derived independently ($N_i = 1$ and $K = N$), then $P = (0.5 + 0.5 \times 0.75)^K = 0.875^K$. If all strains are not independent, the probability that all the N_i strains derived from male i are B/B is $P = 0.5 + 0.5 \times 0.75^{Ni}$. If inbreeding was started from N4 individuals (instead of N3), the value of P for the strain would be $P = 0.5 + 0.5 \times 0.875 = 0.9375$. For a strain derived from N5, $P = 0.5 + 0.5 \times 0.9375 = 0.96875$. If *NA*, *NB*, and *NC* strains were derived from N3, N4, and N5 progeny of the same backcross male, respectively, the probability that all *NA* + *NB* + *NC* strains be homozygous for the B6 allele at a given locus would then be $P = 0.5 + 0.5 \times (0.75)^{NA} \times (0.875)^{NB} \times (0.9375)^{NC}$. The expected fraction of the genome not covered by the set of IRCS is the product of the *P*'s calculated for each contributing backcross male. Theoretically, deriving a single N3 strain from each of 18 N2 males would provide 91% genome coverage. Ninety-six percent genome coverage can be achieved with either three N4 or six N5 strains from each of 18 N2 males.

Consomic strains were developed both at the Institut Pasteur and at the Universidad de la Laguna, Teneriffe (Spain) by backcrossing five or six times SEG/Pas to B6 mice. Chromosome 19 consomic was produced using the STF/Pas strain, another unrelated *Mus spretus* moderately inbred strain. At every generation, mice carrying the chromosome of interest were selected by genotyping four to six microsatellites evenly spread along the chromosome.

Strain nomenclature

The set of IRCS was named BcG. Each strain was named after the number of the fertile N2 male it was derived from, followed by a letter indicating the strain order. For example BcG-122C was the third strain derived from the 122 N2 male. However, for simplicity, it is

often referred to as 122C. Consomic strains were also named BcG followed by the number of the chromosome (e.g. BcG-14 for chromosome 14), with the exception of strain BcF-19 in which chromosome 19 was introgressed from STF/Pas.

Genotyping

IRCS were genotyped three times at different stages of inbreeding. Forty-six strains were genotyped for 95 microsatellites at an average F = 8.7. Fifty strains were later genotyped for 183 microsatellites at an average F = 24.6. All 55 IRCS were finally genotyped for 490 SNPs and 183 microsatellites at an average F = 45.1. The list of markers is available from the web site mentioned below. In order to identify all alleles present in every strain, genotyping was performed on DNA pools prepared from tail clips collected from all breeding pairs or trios (eight to 12 mice per strain). The presence of the two parental alleles in a given sample (scored as "heterozygous") reflected the fact that they were still segregating in the strain.

Primers for microsatellite amplification were purchased either from Research Genetics (Huntsville, AL) or from Genset (Evry, France). Amplification was performed according to the manufacturer's instructions (35 cycles with 55°C or 52°C annealing temperature and 1.5 to 2 mM $MgCl_2$). For each marker, controls included DNAs from B6, SEG/Pas, and 1:1 and 1:20 mixtures of B6 and SEG/Pas, to verify that one B/S heterozygous mouse could be detected among a group of 10 B/B homozygous mice. PCR products were resolved on 4% Nu-Sieve agarose gels stained with ethidium bromide. SNP genotyping was performed at the Centre National de Génotypage (Evry, France) with the Illumina platform.

Hematology

All measurements of blood parameters were made on 60 ± 5 day-old males. Blood was collected by puncture of the orbital sinus and EDTA was immediately added to prevent

clotting. Complete blood counts were determined using a Vet'ABC counter (SCIL, Viernheim, Germany).

Analysis

Statistical analysis was performed using StatView F- 5.0 (SAS Institute Inc., Cary, NC). Groups were compared by one-way ANOVA after checking distributions of values for normality by comparison with a normal distribution with the same mean and standard deviation using a Kolmogorov-Smirnov test. QTL analysis in the F2 cross was performed using R-QTL (BROMAN *et al.* 2003).

A computer program was developed in Turbo-Pascal (Borland, version 7.0) to analyze pedigree data and genotypes, to generate the maps and diagrams shown in this paper, and to create html files. All maps, genotypes and statistics are publicly available at http://www.pasteur.fr/recherche/unites/Gfons/ircs/ircshome.

RESULTS

Establishment of IRCS

All F1 and most backcross males derived from a cross between SEG/Pas and B6 mice are sterile due to hybrid sterility genes (HALDANE 1922). To avoid the deleterious effect of these genes in further generations, the production of IRCS was started from fertile backcross males. Out of 157 N2 males, 18 (11%, approximately 1/8) proved to be fertile, consistent with the previous report of three genetically independent hybrid sterility genes (ELLIOTT *et al.* 2001; GUÉNET *et al.* 1990; PILDER 1997).

In order to achieve the highest possible genome coverage, several strains were initiated from the progeny of each fertile N2 male (Figure 1). We estimated that, with an average of three strains for each of the 18 males, less than 1% of the genome would be missed (except for regions containing hybrid-sterility genes). A total of 118 strains were started, with one (male 49) to nine (male 122) strains per N2 male (average 6.6).

Breeding performances were very poor during the first generations of inbreeding (Figure 2). To avoid even more severe losses, one (and sometimes two) additional backcrosses to B6 were performed. Forty-eight strains underwent one additional backcross. Among those, 14 were unsuccessful due to male sterility. For six other strains, a second backcross was necessary and allowed to rescue the strain. Interestingly, 10 strains required an additional backcross at a late stage of inbreeding (F>10) due to poor breeding, four of which were between F31 and F53. In total, 53% of the strains were lost during inbreeding, with large variations across males. The collection of IRCS currently consists of 55 strains (Figure 3).

Because of poor breeding performance, strict brother-sister mating was sometimes impossible and, in such cases, crosses involved males and females originating from different cages of the strain. This relaxed inbreeding resulted in delayed progression to homozygosity, so that we considered that inbreeding would be achieved after approximately 40 generations.

The number of inbreeding generations varies between 22 and 83 (average 52). Forty-six IRCS have passed F40.

Establishment of ICS

The breeding of ICS was also very difficult and most strains were lost sooner of later during backcrosses. For many chromosomes, mice could not be bred passed the second or third generation. Even for the remaining strains, it was often necessary to split the chromosome in order to get homozygous, viable, and fertile progeny. For chromosomes 6, 13, 14, 16 and 18, only part of the chromosome of SEG/Pas origin could be bred to homozygosity. However, the entire chromosome 19 from STF/Pas could be retained.

Genotyping of IRCS

Genotyping at high density revealed that each strain carries a small number of chromosomal segments of *Mus spretus* origin. Figure 4 shows the example of IRCS 66H with three segments on chromosomes 1, 13 and 18. The boundaries of the each segment were set halfway between neighbor markers with B/B and S/S genotypes. The fraction of the genome covered by the entire set was estimated by adding-up the chromosomal segments of *Mus spretus* origin carried by each strain. For example, eight strains were found to have at least one *Mus spretus* allele –either heterozygous or homozygous – on chromosome 10 (Figure 5). The resulting segments were calculated and superimposed to estimate that approximately 57.9% of this chromosome is covered with IRCS. There was large heterogeneity in genome coverage among chromosomes. No SEG allele was found on chromosome X, probably as a consequence of the selection of fertile N2 males which were hence B6 homozygous. The percentage of SEG genome was less than 12% on chromosomes 5, 8, and 9. By contrast, it

was around 90% for chromosomes 18, and 19, the two smallest chromosomes of the genome. It ranged from 28 to 73% for all other chromosomes.

Maps of the entire genome were established similarly for IRCS and ICS and are presented in Figure 6. An interactive version of these maps can be found on a dedicated web site (http://www.pasteur.fr/recherche/unites/Gfons/ircs/ircshome) which provides an easy access to genotyping data and selection of strains of interest for a particular genomic region. Overall, IRCS and ICS taken separately provide genome coverage of 39.7% and 18.4%, respectively, and 45.6% when combined, out of which 42.9% is fixed as homozygous, and 2.7% was segregating at the time of genotyping. Based on the actual genealogy of the strains, it was expected to be 98% outside of regions containing hybrid-sterility genes.

The average percentage of *Mus spretus* genome per strain was 1.37%, spanning from 0 to 3.79% (Figure 7A). Thirty-seven strains carry less than 1.5% of SEG alleles, and no spretus allele was found in five IRCS (103B, 122E, 135D, 137H, and 142C). Because the set of IRCS is a mixture of strains derived from three, four, or five backcrosses, the average percentage of *Mus spretus* genome was expected to be approximately 8.3%, 6 times higher than what was actually observed.

A total of 173 chromosomal segments of SEG origin were identified, resulting in an average of 3.15 and a maximum of 8 segments per strain (Figure 7B). Interestingly, 11 IRCS carry only one segment, which make them similar to classical congenic strains. Despite the genotyping of over 670 markers, there are still several gaps in the map and some small-sized segments may have been missed, resulting in an underestimation of the total number of segments. Seventy-six segments are identified by only one marker. Based on the available data, the physical size of each segment in Mb was estimated to be 11.7 Mb on average, with a range from 0.5 to 40 Mb (Figure 8).

Since the first two series of genotypes were obtained on strains which were not at the same stage of inbreeding, the relationship between the number of generations of inbreeding and the percentage of SEG genome retained was investigated (Figure 9). Two phases emerged from this analysis. During the first generations of inbreeding, the percentage of spretus genome regularly decreased by approximately 0.3% per generation and reached 1.5% at F15. It remained more or less stable in further generations. This evolution suggests that there was a continuous selection against SEG alleles during the inbreeding process, as long as chromosomal segments were still heterozygous.

Analysis of quantitative traits using IRCS

In order to evaluate the power of IRCS for the genetic dissection of quantitative traits, we analyzed a subset of them for seven hematological parameters, measured on 9 week-old males. Fifteen to 31 mice from 16 IRCS and 14 SEG mice were compared with 28 B6 mice for red blood cell parameters, as well as platelet and white blood cell counts (Table 1 and supplemental Table 1). Four strains did not differ from B6 (120C, 135B, 135E, 157F), while two strains (49A, 137F) showed differences for four of the seven parameters, including hematocrit which was not statistically different between B6 and SEG. Out of 112 comparisons, 21 differences were significant at the 0.05 threshold, eight of which remained significant after Bonferroni correction for multiple testing (threshold = 0.00045).

Strain 66H showed a mean cell volume (MCV) intermediate between that of B6 and SEG, and platelet counts (Plt) higher that than of both parental strains, with a high level of significance (Table 1). This strain carries three SEG segments on chromosomes (Chr) 1, 13, and 18 (Figure 4). An F2 generation was produced between 66H and B6. Seventy-six males were analyzed under the same conditions as the parental strains, and genotyped for

microsatellites marking the three genomic regions. However, no significant association could be observed between either traits and the markers.

The three SEG segments were isolated in congenic strains produced in two generations from an intercross between 66H and B6. These congenic strains were then intercrossed to combine two segments in the same strain. However, the Chr 1 + 13 combination was more difficult to breed and could not be included in the analysis for lack of animals. Congenic and bi-congenic strains were evaluated for MCV and Plt. For both traits, the phenotypic value of the bi-congenic strains could not be anticipated from those of the congenic strains. Indeed, regarding MCV, Chr 18 was similar to B6, while Chr 1 and Chr13 resembled 66H (Figure 10A and Table 2). However, Chr 1 + 18 was not different from B6, and Chr 13 + 18 had a MCV even higher than that of Chr 18 alone. This suggested that Chr 1 and Chr 13 were mostly responsible for the lower MCV observed in 66H, but their individual effect might be abolished in the presence of Chr 18.

Regarding Plt, Chr 13 alone did not seem to have any effect, while Chr 1 and Chr 18 had intermediate values between those of B6 and 66H (Figure 10B and Table 3). Chr 18 was not affected by the presence of Chr 13 in the Chr 13 + 18 bi-congenic, and it seemed that Chr 1 and Chr 18 were able to add up their individual effects in the Chr 1 + 18 bi-congenic, although this observation remains preliminary because of small sample size.

These results show MCV and Plt are controlled by shared genetic regions. They also illustrate the importance of epistasis in the genetic control of these traits. The effect of each region, alone or in combination, is different between the two traits, which suggests that MCV and Plt are controlled by linked but distinct genes.

DISCUSSION

Genetic analysis of complex traits poses special challenges which have hampered so far the characterization of the genes themselves, beyond the identification of QTL-containing regions (CHURCHILL *et al.* 2004; FLINT *et al.* 2005). The development of new resources including large reference populations, and the massive accumulation of phenotypic, high density genotyping, and gene expression data in tens of inbred strains will certainly represent a major instrument towards this goal (WANG *et al.* 2003).

The development of a set of recombinant congenic strains with an interspecific combination of parents aimed at providing a tool complementary to other, existing, resources. Recently, two sets of consomic strains have been successfully developed with the contribution of either *Mus musculus musculus* (strain PWD/Ph (JANSA *et al.* 2005)) or *Mus musculus molossinus* (strain MSM (OKA *et al.* 2007)) wild-derived inbred strains. Crosses with *Mus spretus* have the deserved reputation of being poorly productive, which was the case in the present experiment. Many IRCS failed to breed very early during the inbreeding process, most likely because mice carried allelic combinations which resulted in reduced fitness, failure to thrive, diseases, or sterility. Since the two species have been separated for over 1.5 million years (GUENET and BONHOMME 2003), it is not surprising that genomic incompatibilities have accumulated between unlinked genes. Such epistatic interactions are still compatible with the production of an interspecific backcross, although the small litter size usually observed suggests that there is substantial embryonic lethality.

The strategy used for the development of IRCS proved to be effective. The deleterious effect of hybrid sterility genes was discarded at the first backcross generation. During later generations, relaxed inbreeding imposed by the small number of progeny available to form subsequent breeding pairs has slowed down the progression to full homozygosity, hence providing more chance for unfavorable allelic combinations to be eliminated. To illustrate this

point, let us consider an essential hetero-dimeric complex. It must be assembled from compatible subunits. The genes encoding the two components must have evolved in a concerted fashion so that these components remain compatible throughout evolution. However, when two divergent genomes are mixed and forced to re-assort such as during the development of IRCS, some zygotes will be B6 homozygous for a subunit and SEG homozygous for another one, which may result in the fatal inability to produce a functional complex. All other allelic combinations are compatible with survival because they allow the formation of a functional complex. When genomes become homozygous, both loci must fix alleles of the same origin. Because the SEG contribution to IRCS was, from the beginning, seven times less than the B6 counterpart, it was much more likely that both loci become homozygous for the B6 allele. This is reasonable explanation for the progressive loss of SEG alleles during the first generations of inbreeding.

By comparison, the development of consomic strains forced such essential epistatic interactions between unlinked genes to be disrupted, resulting in the loss of individuals carrying incompatible allelic combinations. As a consequence, only few genomic regions could be retained in consomic strains, especially in the smallest chromosomes (16, 18, and 19).

Deleterious epistatic interactions have been described in very few instances. The best documented example is probably the very strong and consistent transmission ratio distortion observed on chromosome X, in [C57BL/6 × *Mus spretus*] × *Mus spretus* crosses. It was demonstrated that the large excess of *spretus* homozygotes observed for loci located on central chromosome X could be due to the loss of heterozygotes which were not simultaneously heterozygous in a distal region of chromosome 2 (MONTAGUTELLI *et al.* 1996). While such a dramatic situation remains the only observation in interspecific

backcrosses and may be rare, interactions with weaker effects can also have a significant impact in combination with others over several generations.

More than half of the strains were lost during the inbreeding phase, mostly, but not only, during the early generations. Noticeably, four strains stopped breeding after F30 and required an additional backcross. Strain extinction is a common observation during the establishment of recombinant inbred strains, even after full inbreeding (B.A. Taylor, personal communication). Rate of extinction was variable between N2 males, which could reflect differences in the heterozygous regions present in each male. However, no material was collected on these N2 males to perform genotyping.

The average rate of SEG genome in IRCS is 6 times less than expected based on the breeding records and results from a progressive loss of *Mus spretus* alleles occuring during the first 15 generations of inbreeding, when the genome was still largely heterozygous. We observed that in 87% of cases, strains segregating for the two alleles at a particular locus in the first series of genotyping (average F=8.7) had fixed the B6 allele in the second series (average F=24.6). No such genome-wide selection has been observed in recombinant inbred or recombinant congenic strains established from laboratory strains, which showed the expected proportion from the two parental strains (GROOT *et al.* 1992; STASSEN *et al.* 1996). The final genetic make-up of IRCS and ICS suggests that it is hardly feasible to introduce more than 4 to 5% of *Mus spretus* genome in a B6 background, because of the large number of inter-chromosomal epistatic interactions.

The strong reduction of SEG alleles resulted also in that strains derived from the same backcross male were as genetically different as strains derived from independent pedigrees, unlike intraspecific recombinant congenic strains, where strains derived from the same backcross male share on average half of the contribution from the donor strain.

The proportion of SEG genome in IRCS was variable among strains. For five strains, we could not detect any SEG allele, which could be due to either an extreme selection against *Mus spretus* genome, or to the few gaps that still remain between genotyped markers. Since a number of fragments identified are very small in size, it is possible that a few of them were missed. However, it is unlikely that genotyping at higher density will modify the global rate of SEG genome.

Altogether, this set of IRCS covers 39.7% of the mouse genome. This rate is highly variable between chromosomes and varies from 0 to over 90%. Part of this variation may be explained by the early elimination of regions containing hybrid sterility genes (such as for chromosomes X and 9). The two smallest chromosomes were best covered. The variation of genome coverage between chromosomes cannot be explained by their difference in length, hence in gene content. Noticeably, SEG allelic frequency also varied along chromosomes, with the highest frequency observed almost always for loci located to the proximal of distal quarters of chromosome (data not shown). In particular, SEG allele was found to be 13.7% at marker *D6Mit201* on distal chromosome 6. To assess whether this higher than expected frequency could be due to transmission ratio distortion favoring the SEG allele, we set up an intercross between IRCS 137C carrying this allele and B6. Distribution of genotypes did not depart from that expected (data not shown), hence did not provide support for a selective mechanism, although the F2 population (N=102) might have not been large enough to detect weak deviation. Further investigation will be needed to understand the basis for variations in SEG allele frequencies along the genome.

The contribution of *Mus spretus* was clustered, in each IRCS, in a small number of small sized chromosomal segments. These segments were much smaller than those found in classical RCS (GROOT *et al.* 1992; STASSEN *et al.* 1996), due to the combination of relaxed inbreeding and counter-selection of *Mus spretus* alleles. Twelve strains are even congenic

strains in that they carry a single SEG fragment. The small number of fragments has two major advantages. First, if a phenotypic difference is observed between B6 and a given strain, its genetic control is readily attributable to a few regions (or even one is some instances). Second, the role of each fragment can be studied in an F2 between this IRCS and B6, where only 1.3% of the genome segregates, on average, which increases the power to detect epistasis. Alternatively, congenic strains carrying each of the segments can be derived directly from F2 progeny, allowing phenotypic studies on a series of genetically identical, sex- and age-matched individuals. Interactions between QTLs can be studied by combining the segments in bi- or multiple congenic strains.

Inbred strains that are genetically highly divergent are more likely to show phenotypic differences for any trait studied than pairs of laboratory strains. However, interspecific crosses have been criticized as possibly involving too many QTLs for each of them to be detectable. IRCS still capture high rate of polymorphism but it is restricted to small-sized regions. With their reduced proportion of *Mus spretus* genome, it is likely that phenotypic differences be controlled by a small number of QTLs which are amenable to genetic dissection. The high density of polymorphic markers in the *Mus spretus* chromosomal segments will facilitate fine genetic analysis and positional cloning of QTLs. Finally, because the genomes of the two parental strains have diverged for a long time, it is likely that new phenotypes can emerge from the disruption of co-adapted allelic combinations.

The SEG strain was only partially inbred at the time of the first cross with B6. One cannot exclude that different IRCS carrying apparently the same chromosomal region from SEG could actually carry different *Mus spretus* segments with some sequence variation which could confound the interpretation of results. However, high density SNP data has revealed that SEG and STF strains, which originate from Spain and Tunisia, respectively, differ for

only eight of the 660 SNP analyzed (1.2%, data not shown). It is reasonable to assume that genetic polymorphism present within the original SEG colony was even smaller.

QTL mapping using IRCS should not be based on the comparison of phenotypes between the parental strains B6 and SEG for at least three reasons. First, less than half of the potential QTLs can be identified using IRCS and ICS due to partial genome coverage. Second, a difference between B6 and an IRCS may be observed even in the case where B6 and SEG show similar phenotypes. Third, the direction of phenotypic differences between B6 and any IRCS cannot be predicted from the direction observed between B6 and SEG. These last two points have already been reported with RCS and RIS between laboratory strains (GRISEL *et al.* 1997; VAN WEZEL *et al.* 1996). IRCS are best used by assessing the phenotypic trait of interest in all strains of the set by comparison with B6. Any phenotypic difference observed must be controlled by one or more of the chromosomal segments carried by the IRCS which can then be submitted to genetic analysis.

To evaluate the potential of IRCS, we investigated seven hematological traits on 16 IRCS and found eight significant differences in six strains, after Bonferroni correction for multiple testing. This high rate of phenotypic variation between strains likely reflects both the polygenic control of the traits and the high genetic polymorphism of the B6-SEG cross. Similar observations have been made on other traits, related to male fertility (LHOTE *et al.* 2007) and skull morphology (manuscript in preparation). In all cases, it is surprising to note that phenotypes distinct from that of B6 are found despite the low percentage of SEG genome.

Strain 66H was found to differ from B6 for MCV and Plt. None of the three chromosomal regions known to be of SEG origin in 66H showed a significant association with either trait in the F2 cross between 66H and B6. This could be explained by the small size of the F2 cross (N=76), the dominant or recessive mode of inheritance of the SEG alleles, the weak effect of each QTL, and the by complex genetic interactions between the three

regions. This was confirmed by the analysis of the three congenic and two of the three bi-congenic strains. The effect of each region was weak, but significant. Statistically significant epistatic interactions were found and the phenotype of the bi-congenic strains could not be deduced from that of the single congenic strains. For MCV, both Chr 1 and Chr 13 congenic strains appeared to reproduce the phenotype of 66H. Every strain which carried the chromosome 18 segments had a significantly higher MCV than B6, while both chromosome 1 and chromosome 13 congenic strains appeared to reproduce the phenotype of 66H. QTLs for MCV have been reported on chromosomes 1 and 13 by Valdar et al. (VALDAR *et al.* 2006) using heterogeneous stocks. However, their confidence intervals do not overlap the SEG regions of strain 66H and it is therefore likely that they are not identical to the QTLs identified in the present study. Fine mapping of these QTLs is under progress and involves the production of sub-congenic strains, taking advantage of the high density of polymorphic markers.

None of the congenic or bi-congenic strain had a platelet count as high as that of 66H. Chr 1 and 18, separately or combined in the bicongenic strain, conferred a significant increase compared with B6, and are therefore potential targets for QTL identification. We were not able so far to combine the three regions in a single tri-congenic strain to reproduce the phenotype of 66H, as exemplified in the work of Morel et al. (MOREL *et al.* 2000). We cannot rule out that 66H carry a small SEG fragment which has remained so far undetected and which would explain the difference between the chromosome 1 + 18 bi-congenic strain and 66H. QTL for Plt have been described on chromosomes 1 and 18 by Valdar et al. (VALDAR *et al.* 2006) and by Cheung et al. (CHEUNG *et al.* 2004) but in regions absent from 66H.

In the present work, we have demonstrated the feasibility and usefulness of interspecific recombinant congenic strains for the genetic evaluation of complex traits, even for QTLs with small individual effects and complex epistatic interactions. In addition to being a unique and

valuable genetic resource, IRCS provide an experimental example of genomic incompatibilities which develop when two lineages become separated by more than one million years of evolution.

Acknowledgements

We are very grateful to Isabelle Lanctin for highly dedicated and careful breeding of the IRCS and to Stéphanie Voegeling and Murielle Rocancourt for expert assistance with microsatellite genotyping. We thank the Centre National de Génotypage for SNP genotyping.

LITERATURE CITED

ABIOLA, O., J. M. ANGEL, P. AVNER, A. A. BACHMANOV, J. K. BELKNAP *et al.*, 2003 The nature and identification of quantitative trait loci: a community's view. Nat Rev Genet **4:** 911-916.

AVNER, P., L. AMAR, L. DANDOLO and J.-L. GUÉNÉT, 1988 Genetic analysis of the mouse using interspecific crosses. Trends Genet. **4:** 18-23.

BAGOT, S., S. CAMPINO, C. PENHA-GONCALVES, S. PIED, P. A. CAZENAVE *et al.*, 2002 Identification of two cerebral malaria resistance loci using an inbred wild-derived mouse strain. Proc Natl Acad Sci U S A **99:** 9919-9923.

BAILEY, D. W., 1971 Recombinant-inbred strains. An aid to identify linkage and function of histocompatibility and other genes. Transplantation **11:** 325-327.

BAILEY, D. W., 1981 Recombinant inbred strains and bilineal congenic strains, pp. 223-239 in *The mouse in biomedical research*, edited by H. L. FOSTER, J. D. SMALL and J. G. FOX. Academic Press, New York.

BANUS, H. A., H. J. VAN KRANEN, F. R. MOOI, B. HOEBEE, N. J. NAGELKERKE *et al.*, 2005 Genetic control of Bordetella pertussis infection: identification of susceptibility loci using recombinant congenic strains of mice. Infect Immun **73:** 741-747.

BREEN, M., L. DEAKLN, B. MACDONALD, S. MILLER, R. SIBSON *et al.*, 1994 Towards high resolution maps of the mouse and human genomes--a facility for ordering markers to 0.1 cM resolution. Hum. Mol. Genet. **3:** 621-627.

BROMAN, K. W., H. WU, S. SEN and G. A. CHURCHILL, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 889-890.

CHEUNG, C. C., I. C. MARTIN, K. R. ZENGER, J. A. DONALD, P. C. THOMSON *et al.*, 2004 Quantitative trait loci for steady-state platelet count in mice. Mamm Genome **15:** 784-797.

CHURCHILL, G. A., D. C. AIREY, H. ALLAYEE, J. M. ANGEL, A. D. ATTIE *et al.*, 2004 The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet **36:** 1133-1137.

DARVASI, A., A. WEINREB, V. MINKE, J. I. WELLER and M. SOLLER, 1993 Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **134:** 943-951.

DAVIS, R. C., E. E. SCHADT, D. J. SMITH, E. W. HSIEH, A. C. CERVINO *et al.*, 2005 A genome-wide set of congenic mouse strains derived from DBA/2J on a C57BL/6J background. Genomics **86:** 259-270.

DEMANT, P., and A. A. M. HART, 1986 Recombinant congenic strains - a new tool for analyzing genetic traits determined by more than one gene. Immunogenetics **24:** 416-422.

ELLIOTT, R. W., D. R. MILLER, R. S. PEARSALL, C. HOHMAN, Y. ZHANG *et al.*, 2001 Genetic analysis of testis weight and fertility in an interspecies hybrid congenic strain for Chromosome X. Mamm Genome **12:** 45-51.

FLINT, J., W. VALDAR, S. SHIFMAN and R. MOTT, 2005 Strategies for mapping and cloning quantitative trait genes in rodents. Nat Rev Genet **6:** 271-286.

FORTIN, A., L. R. CARDON, M. TAM, E. SKAMENE, M. M. STEVENSON *et al.*, 2001a Identification of a new malaria susceptibility locus (Char4) in recombinant congenic strains of mice. Proc Natl Acad Sci U S A **98:** 10793-10798.

FORTIN, A., E. DIEZ, D. ROCHEFORT, L. LAROCHE, D. MALO *et al.*, 2001b Recombinant congenic strains derived from A/J and C57BL/6J: a tool for genetic dissection of complex traits. Genomics **74:** 21-35.

GOUYA, L., F. MUZEAU, A. M. ROBREAU, P. LETTERON, E. COUCHI *et al.*, 2007 Genetic study of variation in normal mouse iron homeostasis reveals ceruloplasmin as an HFE-hemochromatosis modifier gene. Gastroenterology **132:** 679-686.

GRISEL, J. E., J. K. BELKNAP, L. A. O'TOOLE, M. L. HELMS, C. D. WENGER *et al.*, 1997 Quantitative trait loci affecting methamphetamine responses in BXD recombinant inbred mouse strains. J Neurosci **17:** 745-754.

GROOT, P. C., C. J. MOEN, W. DIETRICH, J. P. STOYE, E. S. LANDER *et al.*, 1992 The recombinant congenic strains for analysis of multigenic traits: genetic composition. Faseb J **6:** 2826-2835.

GUENET, J. L., and F. BONHOMME, 2003 Wild mice: an ever-increasing contribution to a popular mammalian model. Trends Genet **19:** 24-31.

GUÉNET, J. L., and X. MONTAGUTELLI, 1994 The contribution of wild specimens to the establishment of the mouse genetic map, pp. 285-298 in *Genetics in wild mice*, edited by K. MORIWAKI and R. KOMINAMI. Japan Sci. Soc. Press, Tokyo.

GUÉNET, J. L., C. NAGAMINE, D. SIMON-CHAZOTTES, X. MONTAGUTELLI and F. BONHOMME, 1990 *Hst-3*: an X-linked hybrid sterility gene. Genet. Res. **56:** 163-165.

HALDANE, J. B. S., 1922 Sex ratio and unisexual sterility in hybrid animals. J. Genet. **12:** 101-109.

IAKOUBOVA, O. A., C. L. OLSSON, K. M. DAINS, D. A. ROSS, A. ANDALIBI *et al.*, 2001 Genome-tagged mice (GTM): two sets of genome-wide congenic strains. Genomics **74:** 89-104.

IDERAABDULLAH, F. Y., E. DE LA CASA-ESPERON, T. A. BELL, D. A. DETWILER, T. MAGNUSON *et al.*, 2004 Genetic and haplotype diversity among wild-derived mouse inbred strains. Genome Res **14:** 1880-1887.

JANSA, P., P. DIVINA and J. FOREJT, 2005 Construction and characterization of a genomic BAC library for the Mus m. musculus mouse subspecies (PWD/Ph inbred strain). BMC Genomics **6:** 161.

LEE, P. D., B. GE, C. M. GREENWOOD, D. SINNETT, Y. FORTIN *et al.*, 2006 Mapping cis-acting regulatory variation in recombinant congenic strains. Physiol Genomics **25:** 294-302.

LEMAY, A. M., and C. K. HASTON, 2005 Bleomycin-induced pulmonary fibrosis susceptibility genes in AcB/BcA recombinant congenic mice. Physiol Genomics **23:** 54-61.

LHOTE, D., C. SERRES, P. LAISSUE, A. OULMOUDEN, C. ROGEL-GAILLARD *et al.*, 2007 Centimorgan-range one-step mapping of fertility traits using Interspecific Recombinant Congenic Mice. Genetics.

MARTIN, B., G. CHAPOUTHIER and R. MOTTA, 1992a Analysis of B10.D2 recombinant congenic mouse strains shows that audiogenic and beta-CCM-induced seizures depend on different genetic mechanisms. Epilepsia **33:** 11-13.

MARTIN, B., C. MARCHALAND, J. PHILLIPS, G. CHAPOUTHIER, C. SPACH *et al.*, 1992b Recombinant congenic strains of mice from B10.D2 and DBA/2: their contribution to behavior genetic research and application to audiogenic seizures. Behav. Genet. **22:** 685-701.

MASHIMO, T., M. LUCAS, D. SIMON-CHAZOTTES, M.-P. FRENKIEL, X. MONTAGUTELLI *et al.*, 2002 A nonsense mutation in the gene encoding 2'-5'-oligoadenylate synthetase/L1 isoform is associated with West Nile virus susceptibility in laboratory mice. Proc. Natl. Acad. Sci. USA **99:** 11311-11316.

MELANITOU, E., F. JOLY, M. LATHROP, C. BOITARD and P. AVNER, 1998 Evidence for the presence of insulin-dependent diabetes-associated alleles on the distal part of mouse chromosome 6. Genome Res **8:** 608-620.

MOEN, C. J., P. C. GROOT, A. A. HART, M. SNOEK and P. DEMANT, 1996 Fine mapping of colon tumor susceptibility (Scc) genes in the mouse, different from the genes known to be somatically mutated in colon cancer. Proc. Natl. Acad. Sci. USA **93:** 1082-1086.

MOEN, C. J., M. SNOEK, A. A. HART and P. DEMANT, 1992 Scc-1, a novel colon cancer susceptibility gene in the mouse: linkage to CD44 (Ly-24, Pgp-1) on chromosome 2. Oncogene **7:** 563-566.

MOEN, C. J., H. J. STOFFERS, A. A. HART, H. V. WESTERHOFF and P. DEMANT, 1997 Simulation of the distribution of parental strains' genomes in RC strains of mice. Mamm Genome **8:** 884-889.

MOEN, C. J., M. A. VAN DER WALK, M. SNOEK, B. F. VAN ZUTPHEN, O. VON DEIMLING *et al.*, 1991 The recombinant congenic strains - a novel genetic tool applied to the study of colon tumor development in the mouse. Mammal. Genome **1:** 217-227.

MONTAGUTELLI, X., R. TURNER and J. H. NADEAU, 1996 Epistatic control of non-mendelian inheritance in mouse interspecific crosses. Genetics **143:** 1739-1752.

MOREL, L., B. P. CROKER, K. R. BLENMAN, C. MOHAN, G. HUANG *et al.*, 2000 Genetic reconstitution of systemic lupus erythematosus immunopathology with polycongenic murine strains. Proc Natl Acad Sci U S A **97:** 6670-6675.

MOTT, R., and J. FLINT, 2002 Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks. Genetics **160:** 1609-1618.

NEWMAN, T. L., E. TUZUN, V. A. MORRISON, K. E. HAYDEN, M. VENTURA *et al.*, 2005 A genome-wide survey of structural variation between human and chimpanzee. Genome Res **15:** 1344-1356.

OKA, A., T. AOTO, Y. TOTSUKA, R. TAKAHASHI, M. UEDA *et al.*, 2007 Disruption of genetic interaction between two autosomal regions and the X chromosome causes reproductive isolation between mouse strains derived from different subspecies. Genetics **175:** 185-197.

PILDER, S. H., 1997 Identification and linkage mapping of Hst7, a new M. spretus/M. m. domesticus chromosome 17 hybrid sterility locus. Mamm Genome **8:** 290-291.

RHODES, M., R. STRAW, S. FERNANDO, E. EVANS, T. LACEY *et al.*, 1998 A high resolution microsatellite map of the mouse genome. Genome Research **8:** 531-542.

ROY, M. F., N. RIENDEAU, J. C. LOREDO-OSTI and D. MALO, 2006 Complexity in the host response to Salmonella Typhimurium infection in AcB and BcA recombinant congenic strains. Genes Immun **7:** 655-666.

SEBASTIANI, G., V. BLAIS, V. SANCHO, S. N. VOGEL, M. M. STEVENSON *et al.*, 2002 Host immune response to Salmonella enterica serovar Typhimurium infection in mice derived from wild strains. Infect Immun **70:** 1997-2009.

SERREZE, D. V., M. PROCHAZKA, P. C. REIFSNYDER, M. M. BRIDGETT and E. H. LEITER, 1994 Use of recombinant congenic and congenic strains of NOD mice to identify a new insulin-dependant diabetes resistance gene. J. Exp. Med. **180:** 1553-1558.

SINGER, J. B., A. E. HILL, L. C. BURRAGE, K. R. OLSZENS, J. SONG *et al.*, 2004 Genetic dissection of complex traits with chromosome substitution strains of mice. Science **304:** 445-448.

STAELENS, J., B. WIELOCKX, L. PUIMEGE, F. VAN ROY, J. L. GUENET *et al.*, 2002 Hyporesponsiveness of SPRET/Ei mice to lethal shock induced by tumor necrosis factor and implications for a TNF-based antitumor therapy. Proc Natl Acad Sci U S A **99:** 9340-9345.

STASSEN, A. P., P. C. GROOT, J. T. EPPIG and P. DEMANT, 1996 Genetic composition of the recombinant congenic strains. Mamm Genome **7:** 55-58.

TAYLOR, B. A., 1978 Recombinant inbred mice: use in gene mapping, pp. 423-438 in *Origin of inbred mice*, edited by H. C. MORSE. Academic Press, New York.

TURCOTTE, K., J. C. LOREDO-OSTI, P. FORTIN, E. SCHURR, K. MORGAN *et al.*, 2006 Complex genetic control of susceptibility to Mycobacterium bovis (Bacille Calmette-Guerin) infection in wild-derived Mus spretus mice. Genes Immun **7:** 684-687.

VALDAR, W., L. C. SOLBERG, D. GAUGUIER, S. BURNETT, P. KLENERMAN *et al.*, 2006 Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat Genet **38:** 879-887.

VAN WEZEL, T., A. P. STASSEN, C. J. MOEN, A. A. HART, M. A. VAN DER VALK *et al.*, 1996 Gene interaction and single gene effects in colon tumour susceptibility in mice. Nature Genet. **14:** 468-470.

WANG, J., R. W. WILLIAMS and K. F. MANLY, 2003 WebQTL: web-based complex trait analysis. Neuroinformatics **1:** 299-308.

YI, N., A. DIAMENT, S. CHIU, K. KIM, D. B. ALLISON *et al.*, 2004 Characterization of epistasis influencing complex spontaneous obesity in the BSB model. Genetics **167:** 399-409.

FIGURE LEGENDS

Figure 1 - IRCS breeding scheme. SEG/Pas males were bred to B6 to produce F1 females which were crossed to B6 males. Resulting backcross males were selected for fertility and mated with B6 females. From their progeny, breeding pairs were established as a starting point for the development of inbred strains. As soon as possible, each strain was maintained as 4 pairs or trios, with relaxed inbreeding. One to nine breeding pairs could be established from each backcross male. Each strain was named after the number of the backcross male, followed by a letter.

Figure 2 - Percentage of strain extinction during the first generations of inbreeding. A large number of strains did not survive the first 4 generations of inbreeding. Out of 118 breeding pairs, 34 did not yield any progeny. The rate of extinction decreased steadily and the number of strains (given above bars) stabilized after F4, as the result of both the elimination of very deleterious allelic combinations, and the implementation of additional backcrosses to B6 which were able to rescue most infertile strains.

Figure 3 - Schematic representation of the current status of IRCS. Each vertical bar with filled boxes represents one alive IRCS, while open boxes represent extinct strains. Strains derived from the same backcross male are grouped within a horizontal square bracket with the number of the backcross male below. Each box of a vertical bar represents one generation of inbreeding, up to 40. For extinct strains, the number of boxes indicates the number of generations after which the strain stopped breeding. Boxes below the baseline show the additional backcrosses (one or two) that were performed to avoid strain extinction.

Figure 4 - Genetic map of IRCS 66H. IRCS were genotyped for a total of 673 microsatellite and SNP markers represented as ticks along chromosomes, with position in Mb. Regions of B6 origin are depicted in light gray, while SEG segments are drawn in black. Segment boundaries are set halfway between adjacent markers with different genotypes. Strain 66H carries SEG alleles on chromosomes 1, 13 and 18. An interactive version of the maps is available at http://www.pasteur.fr/recherche/unites/Gfons/ircs/ircshome.

Figure 5 - Regions of chromosome 10 covered by IRCS. All strains carrying a fragment of SEG origin on chromosome 10 are represented on the right, as on figure 4. Regions segregating for B6 and SEG alleles are shown in dark grey. The leftmost chromosome combines the contribution of all IRCS. Dashed light grey ticks represent markers with missing genotypes in one or more strains, hence for which there might be undetected SEG alleles. Segment boundaries are set halfway between adjacent markers with different genotypes. It is estimated that 57.9% of this chromosome is covered in the set of IRCS.

Figure 6 - Composite maps of IRCS (A) and ICS (B). Each genetic map summarizes the coverage provided by all IRCS and ICS. Segments in black depict regions for which at least one of the strains carries the SEG allele. Segments in dark grey depict regions for which at least one of the strains was segregating for the SEG allele at the time of genotyping. Segments are interrupted by missing of B6 homozygous genotypes. Dashed light grey ticks represent markers with missing genotypes in one or more strains, hence for which there might be undetected SEG alleles. Segment boundaries are set halfway between adjacent markers with different genotypes. Hatched bar (B) illustrates the STF/Pas origin of the segment carried by the chromosome 19 consomic strain. At present, genome coverage is estimated to be 39.7%

for IRCS, 18.4% for ICS, and 45.6% for both sets combined, out of which 42.9% is fixed as homozygous, and 2.7% was segregating at the time of genotyping

Figure 7 - Distribution of SEG contribution across IRCS. A: Distribution of the percentage of SEG genome. For each strain, the percentage of SEG genome was estimated as the sum of the length of all SEG segments over the total length of the genome (2670 Mb). B : Distribution of the number of SEG segments in each strain. No SEG segment was identified in five strains, while 12 strains carry only one segment and are hence congenic strains (see online maps).

Figure 8 - Distribution of the size of SEG segments in IRCS. Eighty-four percent (146/173) of segments of SEG origin are less than 20 Mb in size (approximately 10 cM), and the average size is 11.7 Mb, similar to what can be expected in regular congenic strains after 25 generations of backcrossing.

Figure 9 - Relationship between the percentage of SEG genome and the number of generations of inbreeding at which IRCS were genotyped. This graph combines data from the first two genotyping series (see text). For each strain, the percentage of SEG genome was estimated as the number of SEG alleles over the total number of genotypes obtained. Symbols reflect the number of backcrosses to B6 performed before inbreeding. Squares : N3; diamonds : N4; triangles : N5. The percentage of SEG genome decreases steadily during the first 20 generations of inbreeding and remains between 1 and 3% in more advanced strains. The dashed line is the regression line calculated from strains between F2 and F20 ($y = 6.43 - 0.28 x$; $r^2 = 0.6$; $p<0.0001$).

Figure 10 - Phenotypes of B6, IRCS 66H, congenic and bi-congenic strains for mean cell volume (A) and platelet counts (B). Bars indicate the mean value (± s.e.m.) measured in groups of age-matched males. Group size and statistical significance of comparisons are shown in Tables 2 and 3.

Table 1. Comparison of SEG/Pas and 16 IRCS with C57BL/6 (N=28 mice) for seven hematological parameters.

Strain	# mice	RBC	Hb	Hct	MCV	MCCH	Plt	WBC
SEG	14	0.0094 B6<SEG	NS	NS	< 10-6 SEG<B6	0.0015 SEG<B6	0.0013 SEG<B6	0.0019 SEG<B6
5A	24	NS	NS	NS	0.00027 SEG<B6<5A	NS	NS	NS
6A	29	NS	NS	NS	NS	NS	$< 10^{-6}$ SEG<B6<6A	NS
6C	25	NS	NS	NS	NS	NS	$7.1\ 10^{-6}$ SEG<B6<6C	NS
49A	20	0.00025 B6<49A<SEG	NS	0.039 SEG<B6<49A	$< 10^{-6}$ SEG<49A<B6	NS	0.046 SEG<B6<49A	NS
66H	20	NS	NS	NS	0.015 SEG<66H<B6	NS	$< 10^{-6}$ SEG<B6<66H	NS
119H	23	NS	NS	NS	NS	NS	0.0016 SEG<B6<119H	NS
120C	24	NS	NS	NS	NS	NS	NS	NS
122C	22	NS	NS	NS	NS	NS	0.0032 SEG<B6<122C	NS
122D	17	NS	NS	NS	$< 10^{-6}$ SEG<B6<122D	NS	0.0045 SEG<B6<122D	NS
122F	15	NS	NS	NS	NS	NS	0.0017 SEG<B6<122F	0.015 SEG<122F<B6
135B	18	NS	NS	NS	NS	NS	NS	NS
135E	22	NS	NS	NS	NS	NS	NS	NS
137E	16	NS	NS	NS	0.014 SEG<B6<137E	NS	NS	NS
137F	25	0.0005 B6<137F<SEG	NS	0.026 SEG<B6<137F	$1.8\ 10^{-6}$ SEG<137F<B6	NS	0.0036 SEG<B6<137F	NS
137G	23	NS	NS	NS	NS	NS	0.013 SEG<B6<137G	NS
157F	31	NS	NS	NS	NS	NS	NS	NS

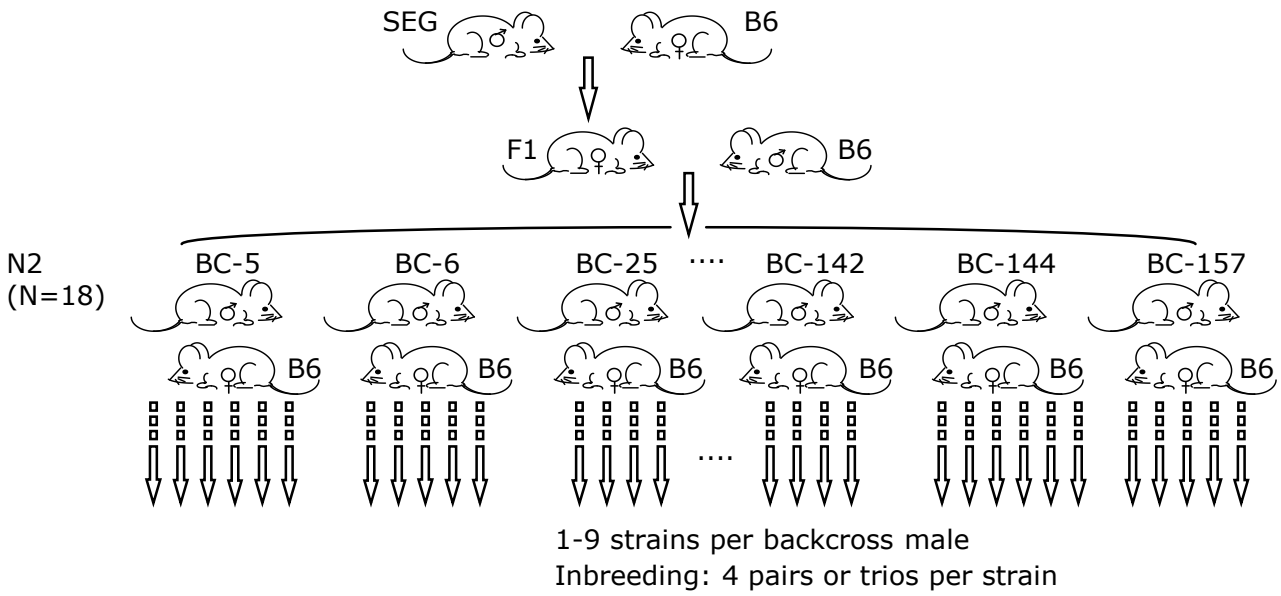
p-value of the comparison between B6 and the strain (NS : p>0.05), with the relative order of the means.

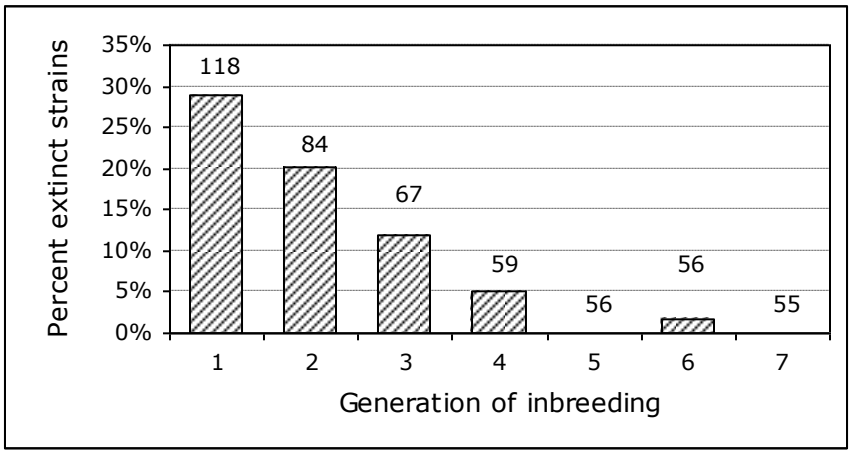
Table 2. Comparison of C57BL/6, congenic, and bi-congenic strains, for MCV. p-values are calculated from Student's t-test.

	# mice	B6	Chr1	Chr13	Chr18	Chr1+18
B6	28					
Chr1	20	0.0005				
Chr13	20	0.0035	NS			
Chr18	27	NS	<0.0001	<0.0001		
Chr1+18	6	NS	0.0018	0.0004	NS	
Chr13+18	17	0.0027	<0.0001	<0.0001	0.04	NS

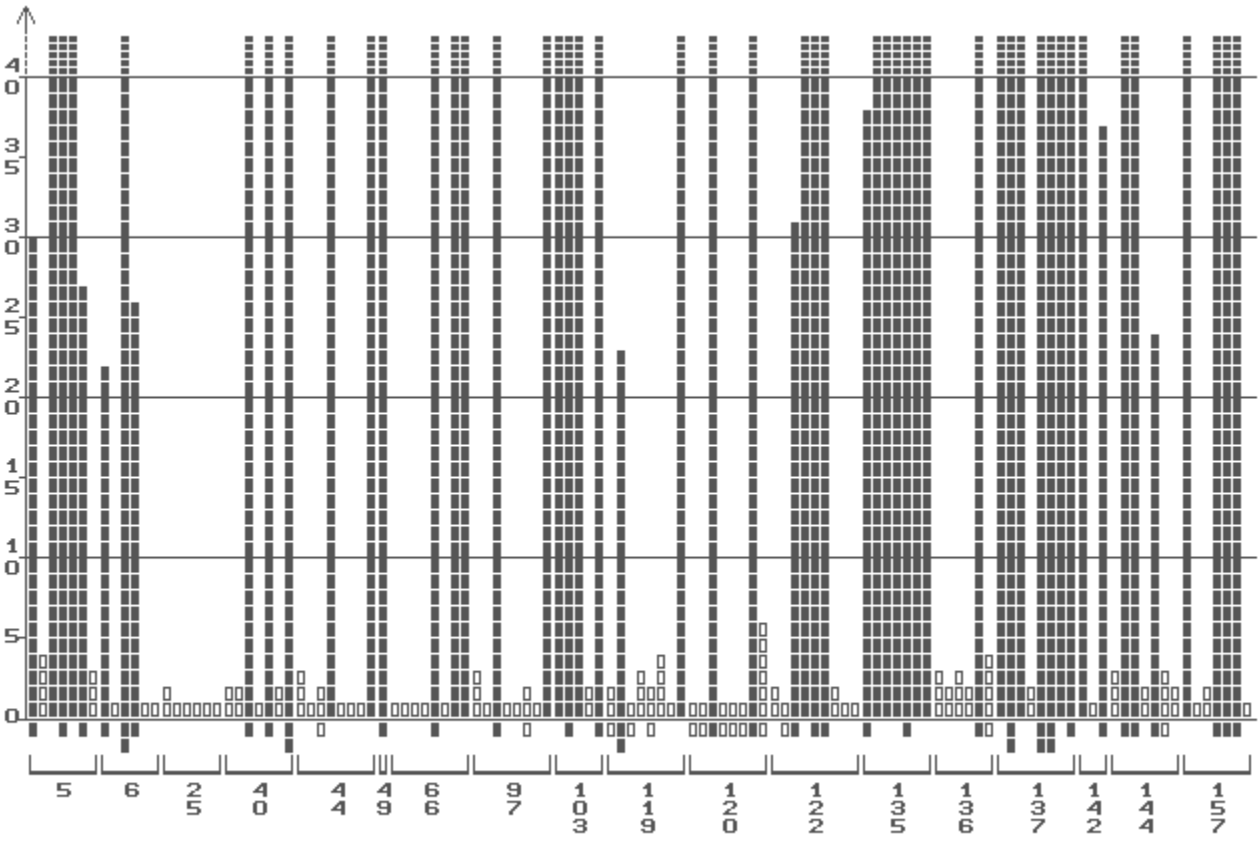
Table 3. Comparison of C57BL/6, congenic, and bi-congenic strains, for platelets. p-values are calculated from Student's t-test.

	# mice	B6	Chr1	Chr13	Chr18	Chr1+18
B6	28					
Chr1	20	0.002				
Chr13	20	NS	0.0001			
Chr18	27	0.018	NS	0.0009		
Chr1+18	6	0.02	NS	0.0046	NS	
Chr13+18	17	0.014	NS	0.0004	NS	NS

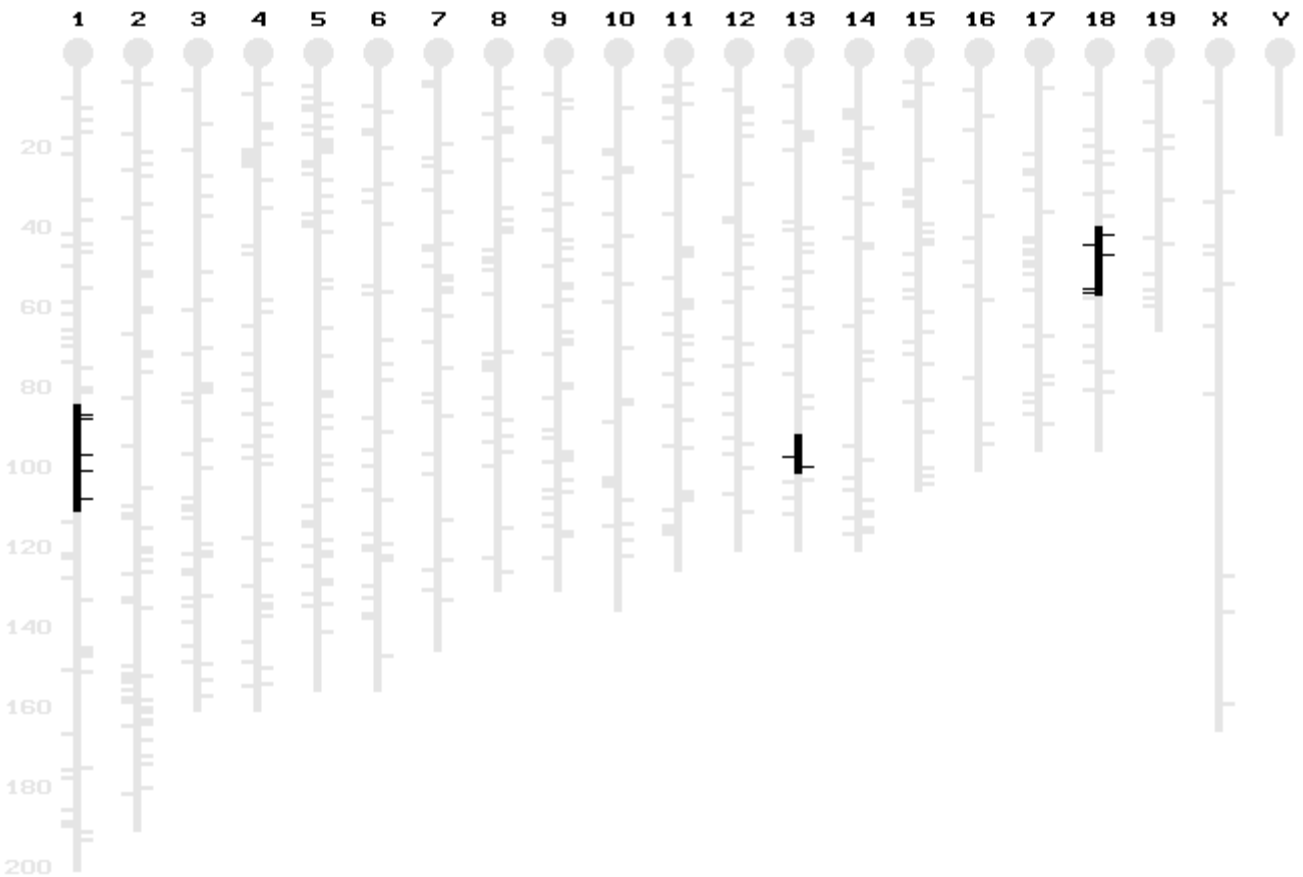

SEG
B6
F1
B6
N2
(N=18)
BC-5
BC-6
BC-25
····
BC-142
BC-144
BC-157
B6
B6
B6
B6
B6
B6
····
1-9 strains per backcross male
Inbreeding: 4 pairs or trios per strain



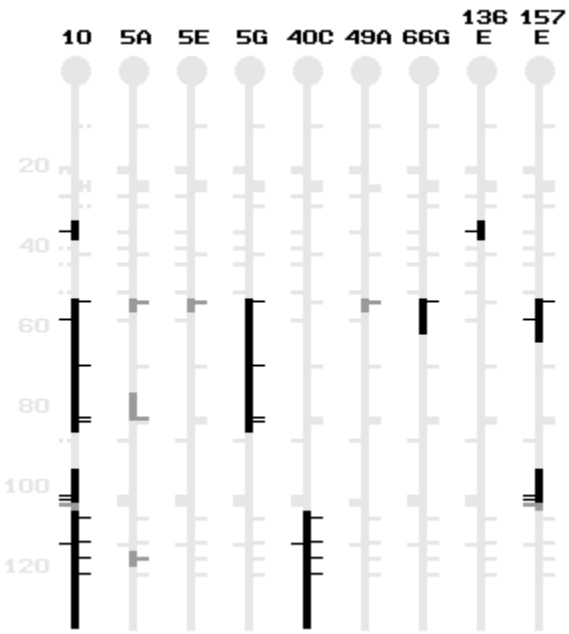
BURGIO et al. - Figure 2

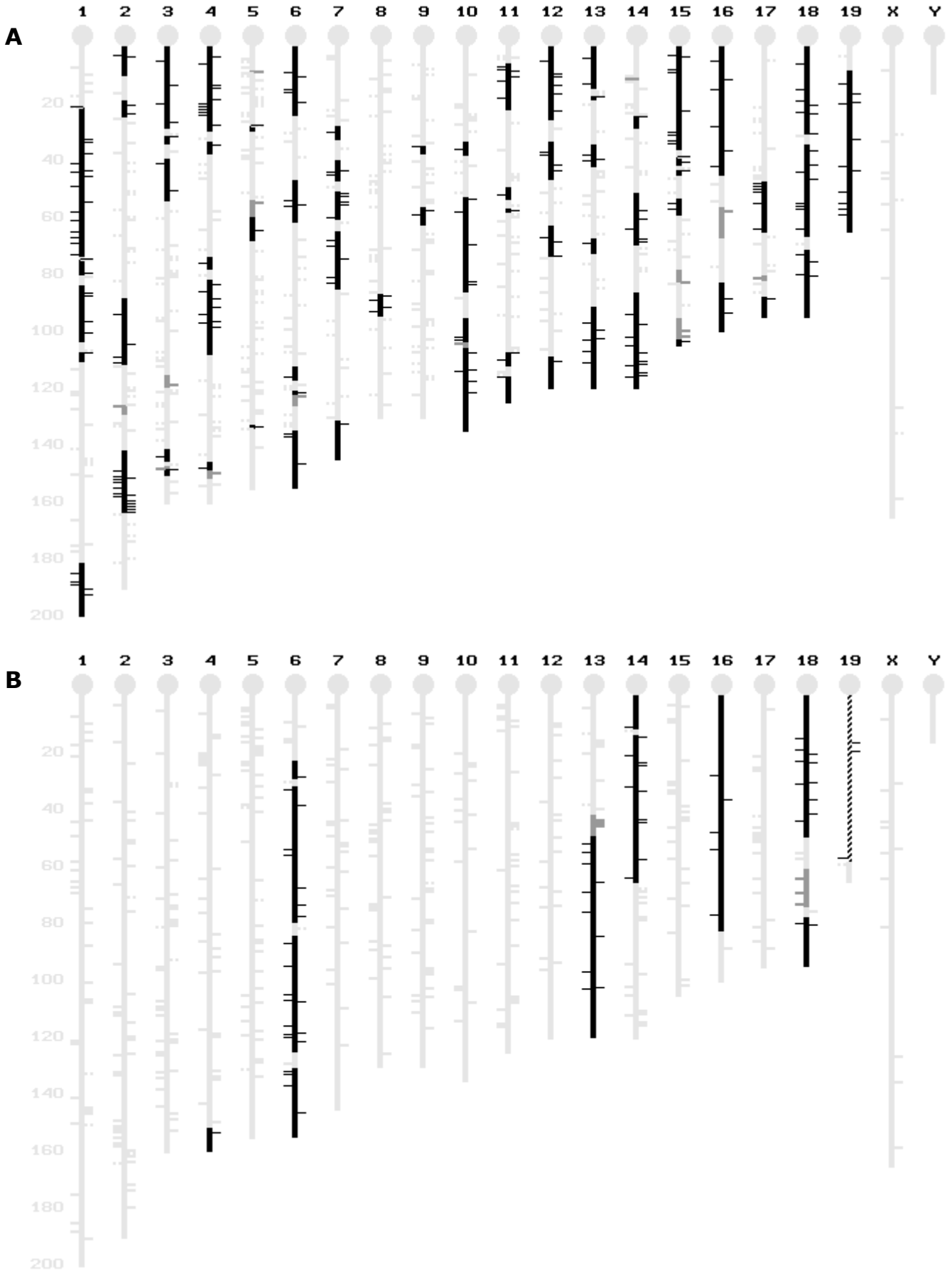


BURGIO et al. - Figure 3



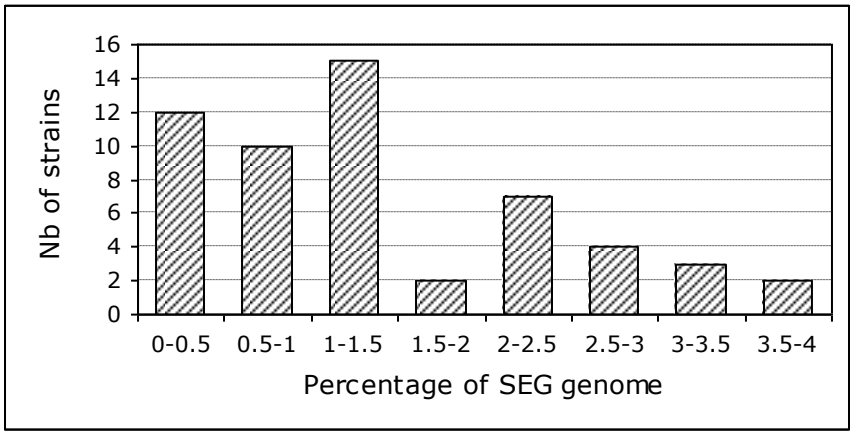
BURGIO et al. - Figure 4


10
5A
5E
5G
40C
49A
66G
136 E
157 E
20
40
60
80
100
120

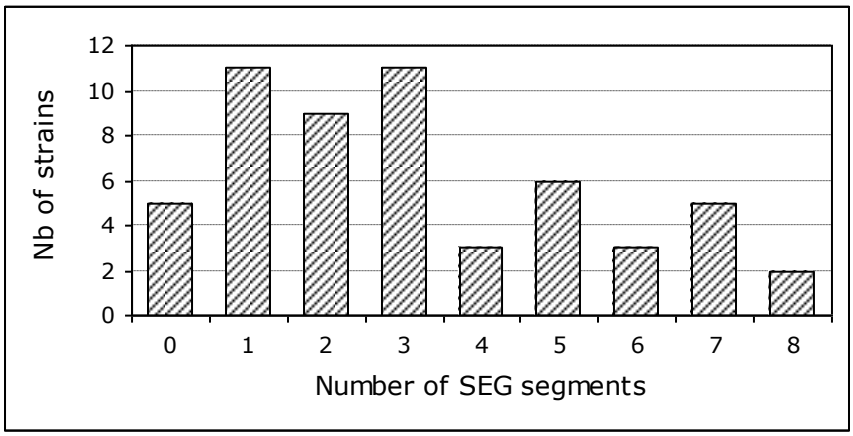


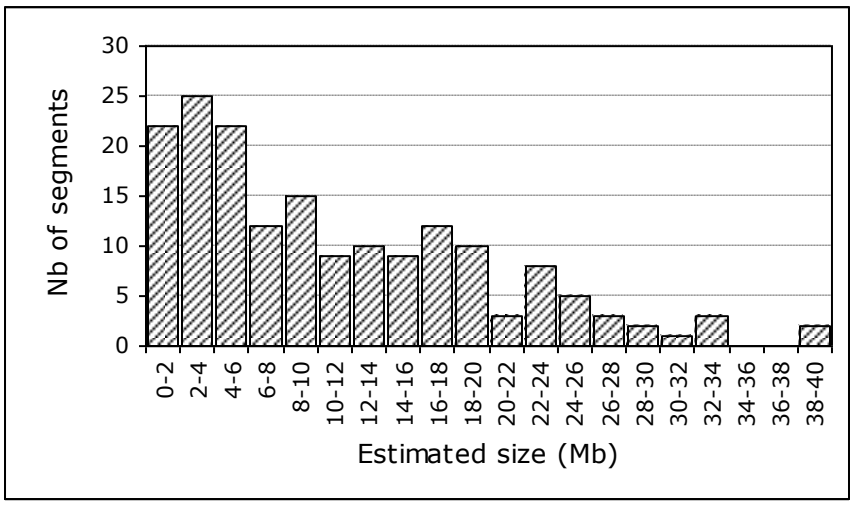
BURGIO et al. - Figure 6

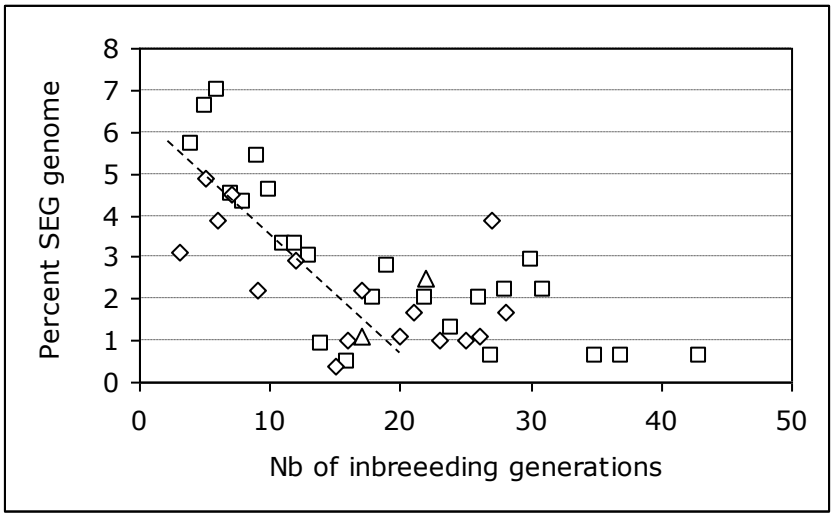
A


Nb of strains
16
14
12
10
8
6
4
2
0
0-0.5
0.5-1
1-1.5
1.5-2
2-2.5
2.5-3
3-3.5
3.5-4
Percentage of SEG genome

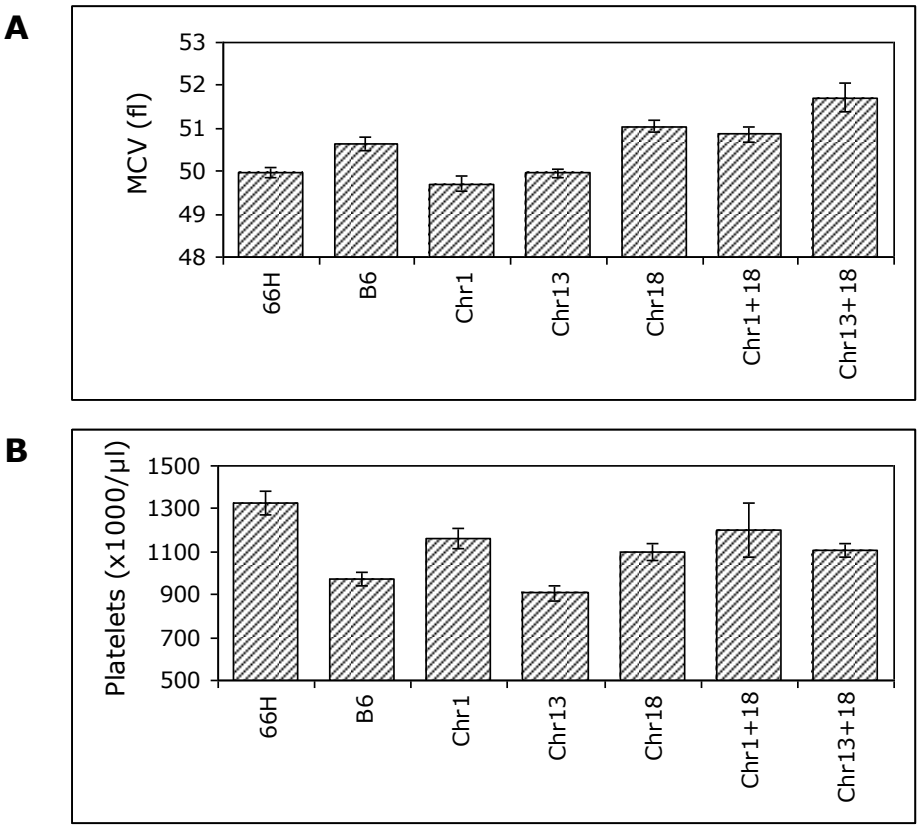
B


Nb of strains
12
10
8
6
4
2
0
0
1
2
3
4
5
6
7
8
Number of SEG segments


Nb of segments
30
25
20
15
10
5
0
0-2
2-4
4-6
6-8
8-10
10-12
12-14
14-16
16-18
18-20
20-22
22-24
24-26
26-28
28-30
30-32
32-34
34-36
36-38
38-40
Estimated size (Mb)



BURGIO et al. - Figure 9


A
MCV (fl)
53
52
51
50
49
48
66H
B6
Chr1
Chr13
Chr18
Chr1+18
Chr13+18
B
Platelets (x1000/µl)
1500
1300
1100
900
700
500
66H
B6
Chr1
Chr13
Chr18
Chr1+18
Chr13+18