GV-SOLAS
Gesellschaft für Versuchstierkunde

Expert information

from the Committee for Genetics and Breeding of Laboratory Animals

Congenic strains by backcross and accelerated backcross ("speed congenics")

Status May 2021

**Authors: Jutta Davidson,
Dirk Wedekind, Kurt Reifenberg,
Reinhart Kluge, Bernhard Aigner**

Contents

1. Introduction ... 3
2. Conventional breeding of congenic strains ... 4
3. "Speed-congenic" strains ... 7
3.1 Choice of markers ... 9
3.2 Marker density ... 9
3.3 Sex chromosomes ... 9
3.4 Number of animals tested per generation ... 10
3.5 Number of backcross generations ... 12
4. Practical problems in the generation of „speed-congenic" strains ... 12
4.1 Non-informative markers of the chosen strains ... 12
4.2 Unequal distribution of the informative markers among the genome ... 12
4.3 Low reproductive performance of the animals ... 12
4.4 Practical examples ... 13
5. Conclusion ... 13
6. Internet addresses for polymorphic genetic markers ... 14
6.1 Mouse ... 14
6.2 Rat ... 14
7. Literature ... 15

1. Introduction

Coisogenic organisms are genetically very similar to each other and differ in only a small part of the genome, usually in only one allele. They are used for the analysis of this allele in the given genetic background. The phenotypic impact of a gene is not only dependent on its specific variants - the different alleles - but also relies on the alleles of many additional genes in the organism. Genes influencing the phenotypic outcome of other genes are named as “modifier genes” (Doetschman 2009). Identity and number of the modifier genes are mostly unknown and may substantially vary for a specific case. E.g., the manifestation of the human height is potentially affected by many thousands of genomic loci (Kaiser 2020).

The analysis of a specific allele on different genetic backgrounds can be carried out by the generation of congenic strains which harbour the identical allele of interest on different genetic backgrounds. For animal research, a great number of inbred strains is available for laboratory rodent species with a standardized and defined genetic background, for to analyse the phenotypic impact of specific alleles on different genetic backgrounds. Two classical inbred strains of laboratory mice of the species *Mus musculus* functionally differ in more than 1,000 of the ca. 20,000 protein coding genes (= approx. 5 - 10 %; Keane et al. 2011, Vanden Berghe et al. 2015). The functional difference between two inbred strains in further genomic loci such as loci coding for non-coding regulatory RNAs or regulatory binding sites (e.g., promoter, enhancer, silencer, etc.) is so far unknown. Thus, the functional difference in these genomic elements between two inbred strains may also be different compared to the situation for protein coding genes.

In addition, specific alleles or transgenic loci are combined from two or more source lines into novel animal models (e.g., by using two-piece transgenic systems like the *Cre/loxP* system). Here, it may be necessary to harmonize the genetic background of the source strains or to standardize the genetic background of the novel line by backcross to a specific inbred strain.

The first congenic strains of laboratory mice were generated around 1920 by George Snell for the research on the rejection of tumour implantations (Silver 1995). Congenic rat strains were mainly developed in the years around 1970 and 1980 in Prague, Cambridge, Hannover und Göttingen to examine the function and structure of the transplantation antigens (MHC; Hedrich 2000).

The generation of congenic strains by conventional breeding techniques in its classical form with many backcrosses of the respective allele into the new genetic background lasts several years. It can be accelerated and optimized by the use of molecular genetic techniques (“speed-congenic” strains). Both the breeding techniques and the genome of the animals derived thereof are described in this manuscript.

Alternatively, genetic engineering techniques may produce organisms which harbour the identical genetic attributes as congenic strains, i.e., the identical allele of interest on different genetic backgrounds. By using sequence-specific nucleases like predominantly CRISPR-Cas9, a specific allele can be principally introduced in the identical way in more than one individual with reasonable cost of time and work. Thus, at least two newly produced individuals harbour the identical DNA sequence of a novel allele in the same locus of the genome. By successfully using two different inbred strains in such a project (if the animals are vital and fertile), the source strain and the novel established strain are coisogenic strains for each inbred

strain, whereas the two newly produced strains harbour the identical genetic attributes as congenic strains (i.e. the identical DNA sequence of a novel allele on the same genomic locus, on different genetic backgrounds) without having to carry out the time-consuming conventional breeding process. Compared to the generation of congenic strains by conventional breeding methods, the disadvantage is the high, but reasonable cost of work at the beginning of the project. However, the two main advantages are the high time-saving effect and the absence of genetic contaminations (see 2.). Potentially caused undesired, unlinked mutations due to the use of the respective genetic engineering techniques ("off-target effects") may be removed from the population by backcross of the allele of interest to the respective genetic background used.

2. Conventional breeding of congenic strains

The generation of congenic strains is carried out by the backcross of the respective allele from the donor strain to the recipient strain (Fig. 1). In principle, more than one allele can be backcrossed at the same time. An inbred strain is usually used as novel recipient strain. Phylogenetic trees of inbred strains of mice (Beck et al. 2000, Petkov et al. 2004) and rats (Canzian 1997, STAR Consortium 2008) are published. The designation of the backcross generations is: F1, N2, N3, and so on. Selection of the allele of interest is carried out in every generation usually on the genetic basis by the use of PCR techniques. This can also be done in the presence of recessive alleles without having to carry out intermediate breeding steps in the backcross protocol.

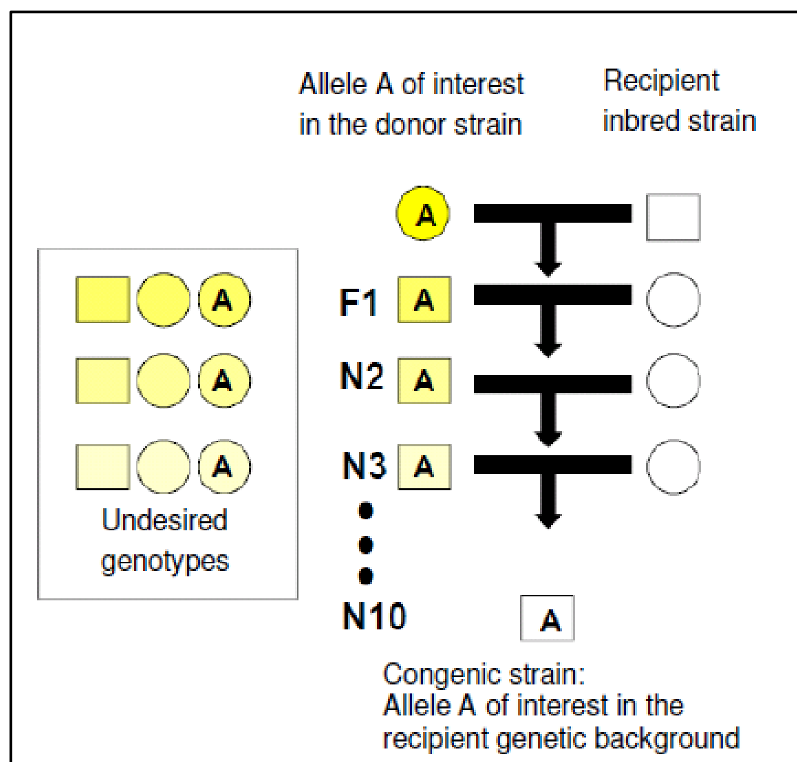

Fig. 1. Breeding scheme of the generation of congenic strains by the backcross of the allele of interest "A" from the donor strain to the recipient strain. Male and female animals are depicted by squares and circles, respectively. Initially, a female carrier of the allele "A" will be mated to a male of the recipient inbred strain. All male offspring of the mating carry the Y chromosome of the recipient strain. In all following generations male carriers of the allele "A" will be selected and mated to females of the recipient strain. This leads to the implementation of the presence of the X chromosome and the mitochondrial DNA of the recipient strain in the generation N2 of the new congenic strain.

The conventional breeding protocol for the generation of congenic strains includes 10 successive backcrosses (F1, N2 – N10) of the donor strain to the recipient strain. The mean generation interval of mice is about 10 - 14 weeks which results in a time period of 2 - 3 years for the generation of conventional congenic strains. This may severely delay the further research on these animals.

The time period which is necessary for the generation of conventional congenic strains may be shortened by the use of artificial reproductive techniques like hormonal induction of the ovulation (super-ovulation) or use of sperm precursor cells (Behringer 1998, Ogonuki et al. 2009, Landa et al. 2010, Behringer et al. 2014). The congenic strain will be maintained by mating of homozygous carriers of the allele of interest or by mating heterozygous / hemizygous carriers with wild-type littermates (Silver 1995).

During the backcross process genomic fragments of the donor strain will be successively replaced by the genome of the recipient strain. For the determination of the genetic contamination of a congenic strain by genomic fragments of the donor strain two different parts have to be considered, i.e., the genomic fragment chromosomally linked to the selected locus, and unlinked genomic fragments (Fig. 2). For the genomic fragments without linkage to the selected locus, the genetic contamination of the donor strain on average decreases by 50 % for each additional backcross generation “n”. The length of the undesired donor strain genome in the diploid genome of the congenic strain (the fraction of donor strain alleles on the total number of alleles) is calculated by the formula 0.5 n. For n = 10, the calculation is 0.001 (= 0.1 %). The genetic contamination of a congenic strain with donor strain genome fragments may be also calculated for the haploid genome which describes the fraction of heterozygous (one allele each of the donor strain and the recipient strain) unlinked loci in the genome of the congenic strain. This contamination is calculated by the formula 0.5 n-1 thereby resulting in the value of 0.002 (= 0.2 %) for n = 10. In this case, on average 40 of the ca. 20,000 protein coding mammalian genes are affected. The functional difference of donor strain and recipient strain for these loci is approx. 5 - 10 % (Vanden Berghe et al. 2015), therefore, particular loci of the genetic contamination in the genome of the congenic recipient strain do not harbour the correct functional allele. The functional difference between two inbred strains in further genomic loci such as loci coding for non-coding regulatory RNAs or regulatory binding sites (e.g., promoter, enhancer, silencer, etc.) is so far unknown. The alleles of the recipient strain will be fixed in half of the heterozygous loci during the ongoing breeding process of the congenic strain via inbreeding. In practice, the genetic contamination with unlinked loci may be higher than calculated in the congenic strain due to unintentional and/or technically immanent selection during the breeding process (e.g., vitality or fertility of the carrier animal only in combination with alleles of the donor strain which are not present in the recipient strain) (Berry und Cutler Linder 2007). Additional backcross generations will lead to the further decrease of the genetic contamination (Fig. 2). The generation of several independent congenic strains for the allele of interest on the identical genome of the recipient strain may be an additional option to exclude the potential influence on the phenotype of congenic strains by unlinked genome fragments of the donor strain (Armstrong et al. 2006).

However, the contamination of the congenic strain with donor strain genome fragments which are linked to the selected locus, has much more impact (Fig. 2). The selected allele will be not transferred alone but combined with a chromosomal fragment of considerable length from the donor strain to the recipient strain during the backcross process. This fragment is named as

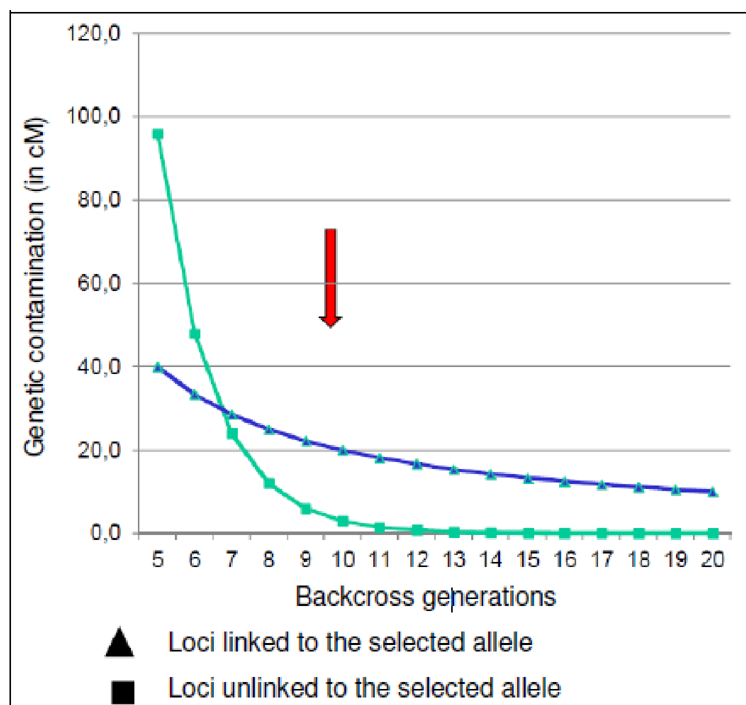

Fig. 2. Contamination of a congenic strain by genomic fragments of the donor strain. The contamination of the linked (triangles) and unlinked (squares) donor strain genome fragments is depicted in Centi-Morgan (cM) for the generations N5 to N20. In generation N10 (red arrow), the contamination of the linked donor strain genome fragments is 20 cM, which refers to approx. 1.3 % of the haploid genome. The contamination of the unlinked donor strain genome fragments in generation N10 is about 0.2 % of the haploid genome referring to ca. 3 cM. Thus, the genetic contamination of a congenic strain is mostly caused by the linked donor strain fragments. The figure is delivered from the University Mainz. [Centi-Morgan is the genetic measurement unit of the recombination frequency between two loci. A distance of 1 cM between two genomic loci refers to the appearance of 1 crossing-over event in 100 meioses. 1 cM refers to about 2 megabase pairs (Mb) in the mouse and rat genome. The length of the haploid genome of mouse and rat is approx. 3,000 Mb.]

"congenic interval" or "differential chromosomal interval". The length of the chromosomal fragments (in cM, for definition see Fig. 2) which will be transferred to the recipient strain by the linkage to the selected allele, is calculated with the formula 200 / n, starting from the backcross generation N5. After 10 backcross generations (N10), the fragment length is approx. 20 cM. As the mouse genome has a length of more than 1,500 cM, the linked genetic contamination results in about 1.3 % of the haploid mouse genome. The subsequent propagation of the congenic strain by inbreeding results in the fixation of the donor strain genome fragment, which is linked to the selected allele, on both autosomes. On the other hand, breeding of heterozygous or hemizygous carrier animals to wild-type littermates for the further propagation of the congenic strain results in the presence of the linked donor strain fragment on only one autosome, i.e., the contamination is 0.66 % of the diploid genome.

1 % of the mammalian genome refers to on average ca. 200 affected protein coding genes. The functional difference between donor strain and recipient strain again is approx. 5 - 10 % of these loci (Vanden Berghe et al. 2015). The functional difference between two inbred strains in further genomic loci such as loci coding for non-coding regulatory RNAs or regulatory binding sites (e.g., promoter, enhancer, silencer, etc.) is so far unknown.

Contaminations of the congenic strain by residual genomic fragments of the donor strain may severely affect the phenotype (Silver 1995, Lusis et al. 2007, Kenneth et al. 2012, Vanden Berghe et al. 2015). This is mostly caused by the contamination of donor strain genome fragments which are linked to the selected locus due to the higher fragment length and the potential interference of neighbouring gene loci in the same functional domains, e.g., in the case of multi-gene families (https://www.ensembl.org/index.html) (Fig. 2).

The nomenclature of congenic strains includes the names of the recipient strain and the donor strain as well as the selected allele. The current recommendations for the nomenclature are published in the internet (http://www.informatics.jax.org/mgihome/nomen/index.shtml; see “Guidelines for nomenclature of mouse and rat strains”). The genetic drift of inbred strains due to the spontaneous mutation rate may be controlled by performing cryo-preservation of embryos (Taft et al. 2006, Uchimura et al. 2015).

For congenic strains which are bred by using substrains of the inbred strain 129 (formerly usually used as genetic background of embryonic stem cells for the generation of homologous recombinants) as donor strain and C57BL/6J or C57BL/6NJ as recipient strain, an internet program is available for the search of potential functional differences in the area of the linked genetic contamination or of further genomic loci of interest (http://me-pamufind-it.org; Vanden Berghe et al. 2015).

In practice of biomedical research, results of congenic strains established with breeding regimens of less than 10 backcross generations and without additional genetic analyses are published. As shown above, a higher degree of genetic contaminations is present in these cases.

3. “Speed-congenic” strains

Speed rate and efficiency in the generation of a congenic strain may be markedly increased by the use of a marker assisted selection protocol (MASP). The selection of suitable animals for the breeding of the next generation is hereby not limited to the presence of the allele of interest but includes the additional selection for maximizing the recipient strain genome fragments in the novel strain.

The selection of the optimal animals for breeding the next generation out of the pool of all offspring consists in the genome-wide analysis of genetic markers which show polymorphisms between the two inbred strains involved in the generation of the congenic strain. Microsatellites (STR, short tandem repeats, often “MIT markers”) or SNPs (single nucleotide polymorphisms) are used for this purpose. Microsatellites are direct repeats of small units consisting of two, three or four nucleotides which show alleles of different length. Fragment length polymorphisms (SSLP, simple sequence length polymorphisms) may be detected by using PCR amplification. SNPs are polymorphisms of a single nucleotide at a defined genome locus which can be identified by using methods like RFLP (restriction fragment length polymorphism) or sequencing of the respective PCR product or allele-specific PCR. Compared to microsatellites, SNPs usually are dimorph, i.e., only two alleles exist for a specific SNP. They are suitable for the use in automatized high-throughput analyses (e.g., Yang et al. 2009, Morgan et al. 2016).

The genome of the generation F1 (= F1 hybrids when using inbred strains both for the donor strain and recipient strain) consists of half of the genome from both the donor strain and the

recipient strain. The use of a MASP usually starts in the generation N2 independently of the degree of homozygosity in the donor strain and the recipient strain, as well as of the degree of genetic homology in the donor strain and the recipient strain.

A MASP is often used to maximize only the fraction of the recipient strain in the genomic loci without linkage to the selected allele. In practice, a value of ca. 99.9 % can be obtained for these unlinked genomic loci of the recipient strain already after 5 - 7 backcross generations (Tab. 1). The main breeding progress hereby appears in the first backcross generations. Thus, the time period for the generation of a congenic strain by accelerated backcross usually is ca. 1.5 years for the mouse (Markel et al. 1997, Wakeland et al. 1997, Weil et al. 1997, Visscher 1999, Wong 2002, Armstrong et al. 2006). The use of artificial reproductive techniques may further decrease the time expense (e.g., Ogonuki et al.2009).

Backcross generation	Genome of the recipient strain (%)	
	without MASP	with MASP
N1 (=F1)	50.0	50.0
N2	75.0	80.8
N3	87.5	94.0
N4	93.8	99.0
N5	96.9	99.9
N6	98.4	
N7	99.2	
N8	99.6	
N9	99.8	
N10	99.9	

Tab. 1. Impact of a marker assisted selection protocol (MASP) on the fraction of the diploid genome of a congenic strain derived from the recipient strain which shows no linkage to the selected allele. The values without the use of a MASP refer to the theoretical average in the population and can be calculated by the formula 1-0.5 n. The values including the use of a MASP were calculated for the analysis of 20 animals carrying the allele of interest for each generation and subsequently verified experimentally (Markel et al. 1997).

A MASP should also be used to minimize the large contamination of the chromosomal fragment which is linked to the selected allele (Fig. 2). Without the use of a MASP for these genomic fractions, the length (in cM) of the chromosomal fragment of the donor strain which will be transferred by the linkage to the selected allele to the congenic strain, is again calculated with the formula 200 / n, starting from the backcross generation N5. The analysis of a great number of animals by using the respective genetic markers is necessary to considerably reduce the linked donor strain genome fragment. This may be combined with the accelerated backcross project or carried out after the generation of the congenic strain. In congenic strains harbouring 20 cM of a genetic contamination of the donor strain around the selected allele, it is estimated that this linked genetic contamination may be reduced to 5 cM in two additional breeding steps, if 50 animals carrying the allele of interest are analysed for each generation in the selection process for each the 5' and 3' end of the congenic interval (Silver 1995).

In view of the cost-benefit calculation, following main points have to be considered when carrying out an accelerated backcross project: choice of markers, marker density, differentiation between heterosomes (sex chromosomes) and autosomes, number of animals tested per generation, as well as number of backcross generations.

3.1 Choice of markers

The polymorphic markers should equally cover the whole genome except of the sex chromosomes (see 3.3). Several polymorphic markers per chromosome have to be examined to identify the crossing-over events during the meiosis. For the choice of polymorphic markers, several databases for mouse and rat can be used (see 6. internet addresses), and marker sets are published for many inbred strains (mouse: e.g., Schalkwyk et al. 1999, Witmer et al. 2003, Szatkiewicz et al. 2008, Cox et al. 2009; rat: e.g., Bryda und Riley 2008, STAR Consortium 2008). In addition, commercial services offer tests with established marker sets as well as the individual compilation and test of marker sets.

3.2 Marker density

A marker density with a distance of less than 20 cM is recommended. The complete mouse genome covers more than 1,500 cM; therefore, a marker set of at least 80 markers equally distributed over the whole genome should be used. Computer simulations using mean marker distances of either 10 cM or 25 cM revealed a comparable low risk for the appearance of duplicate crossing-over events leading to genetic contaminations with donor strain DNA (Wakeland et al. 1997). However, another publication described a higher risk for the appearance of unrecognized genome fragments of the donor strain and recommends a marker density with a distance of 10 cM (Armstrong et al. 2006).

Genome fragments of the recipient strain will be successively fixed during the backcross generations. There is no principal need to test these genomic fragments with markers in the subsequent backcross generations. Thus, a decision has to be made for every further generation whether the number of markers and costs should be reduced, or novel markers should be additionally examined (e.g., for the reduction of the linked genetic contamination of the donor strain). A genome-wide genotypic end control of the novel speed-congenic strain is recommended at the end of the backcross project.

3.3 Sex chromosomes

In the case that the allele of interest is located on an autosome, the sex chromosomes of the recipient strain will be fixed in the generations F1 and N2 if the breeding scheme shown in Fig. 1 is applied. There is no further necessity in the use of specific markers for the sex chromosomes.

In the case that the allele of interest is located on the Y chromosome, male carrier animals have to be used already in the first backcross generation. The X chromosome of the recipient strain will be fixed already in generation F1, with no further necessity in the use of specific markers for the sex chromosomes. The selection for the allele of interest is done by using male offspring for the backcross process because all male animals in all generations are carriers of the allele of interest. However, the Y chromosome of the novel congenic strain completely consists of donor strain genome.

In the case that the allele of interest is located on the X chromosome, the use of female carrier animals is recommended in all backcross generations. This will guarantee the possibility of the appearance of recombination events in every generation leading to the maximal reduction of the linked genomic fragment of the donor strain in the novel congenic strain. Here, specific markers for the X chromosome should be used.

3.4 Number of animals tested per generation

The fraction of the recipient strain genome in the offspring of the backcross matings follows a normal distribution. Thus, random selection of carrier animals will result to a high probability in individuals which carry approximately the theoretical mean fraction of recipient strain genome. Animals with higher or lower fractions will comparably appear to a lower probability. The increase of the number of animals tested with the marker assisted selection protocol (MASP) will increase the probability to discover an individual with a relatively high genomic fraction of the recipient strain. Usually, about 10 - 20 animals carrying the allele of interest will be examined with the MASP, and the 2 - 4 individuals with the highest fraction of the recipient strain genome will be chosen for the next backcross (Fig. 3). The analysis of more than 30 - 50 animals will not result in a significant advantage (Markel et al. 1997). Usually, male carriers are used in the matings with several females of the recipient strain to maximize the number of offspring (Fig. 1). Selection and backcross of more than one carrier animal may prevent the appearance of potential fertility problems.

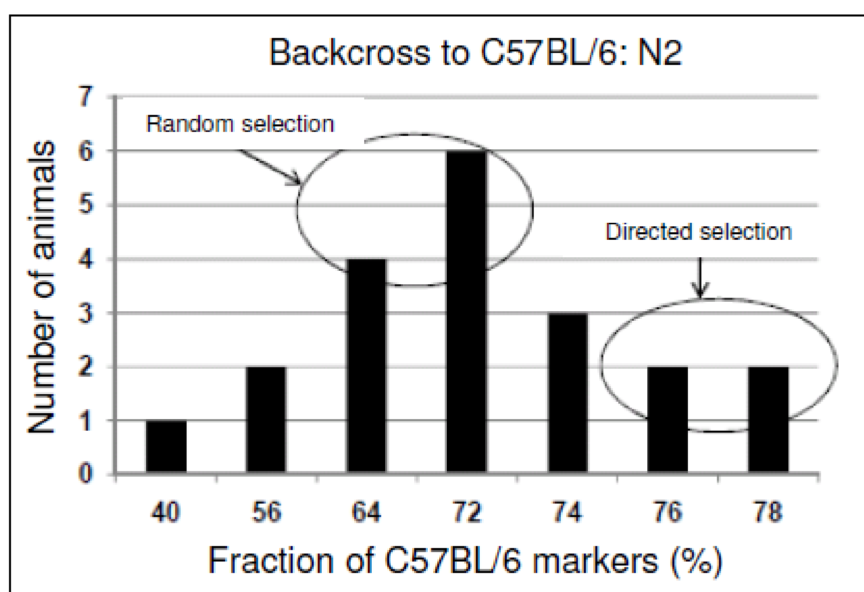

Fig. 3. Practical example of a backcross project using the C57BL/6 inbred strain as recipient strain; backcross generation N2. In total, 20 animals of the generation N2 were examined by the use of a marker assisted selection protocol (MASP). The number of animals with their different genomic fraction of the recipient strain is shown. The fraction of the recipient strain was calculated based on the diploid genome of the animals. The theoretical mean value is calculated by the formula 1-0.5 n (n = backcross generation) and is 75 % for the generation N2. The figure shows that the genomic fraction of the recipient strain in the offspring follows a normal distribution around the theoretically calculated average value. Random choice of carrier animals for breeding the next backcross generation may result with high probability in the selection of individuals harbouring a medium genome fraction of the recipient strain. The directed selection by carrying out a MASP will specifically identify the animals harbouring the highest amounts of recipient strain genome.

The selection of the optimal individuals for the breeding of the next backcross generation may be done according to the highest number of complete recipient strain chromosomes or to the highest total fraction of recipient strain genome. Computer simulations resulted in a similar progress of the success of the backcross project in both cases (Weil et al. 1997). Tab. 2 shows a practical example for the selection by using the highest total fraction of recipient strain genome in the project. In addition, linked loci of the congenic interval and/or project-specific unlinked loci may be examined.

Marker	Chromo-somal position (cM)	N2-1	N2-2	N2-3	N2-4	N2-5	N2-6	N2-7	N2-8	N2-9	N2-10
Chr1-1	10	0.5	1.0	0.5	1.0	1.0	1.0	0.5	1.0	1.0	0.5
Chr1-2	21	0.5	1.0	0.5	1.0	0.5	1.0	0.5	1.0	0.5	0.5
Chr1-3	37	0.5	1.0	0.5	1.0	0.5	1.0	1.0	1.0	0.5	0.5
Chr1-4	47	0.5	1.0	0.5	1.0	0.5	1.0	1.0	1.0	0.5	1.0
Chr1-5	54	0.5	1.0	0.5	1.0	0.5	1.0	1.0	1.0	0.5	1.0
Chr1-6	67	0.5	1.0	0.5	1.0	0.5	1.0	1.0	1.0	0.5	1.0
Chr1-7	94	0.5	1.0	0.5	1.0	0.5	1.0	1.0	1.0	0.5	1.0
Chr1-8	106	0.5	1.0	0.5	1.0	1.0	1.0	0.5	1.0	0.5	1.0
Chr1-1	15	0.5	1.0	0.5	1.0	0.5	0.5	1.0	0.5	0.5	0.5
Chr1-2	34	0.5	1.0	0.5	1.0	0.5	0.5	1.0	0.5	0.5	0.5
Chr1-3	47	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Chr1-4	52	0.5	0.5	0.5	0.5	0.5	1.0	0.5	0.5	0.5	1.0
Chr1-5	58	0.5	0.5	0.5	0.5	0.5	1.0	0.5	1.0	1.0	1.0
C57BL/6 genome (%)		**71.6**	**82.7**	**70.9**	**77.3**	**75.0**	**87.4**	**77.3**	**71.8**	**84.1**	**79.8**
Mean of the C57BL/6 genome (%)		76.9									

Tab. 2. A section (showing only chromosomes 1 and 19) of the results obtained by using a marker assisted selection protocol (MASP). The backcross was carried out using the C57BL/6 inbred strain as recipient strain. The results of the MASP of 10 animals tested (N2-1 to N2-10) of the backcross generation N2 for 8 markers on chromosome 1 (Chr1-1 to Chr1-8, grey) and 5 markers on chromosome 19 (Chr19-1 to Chr19-5, blue) are shown (the results for the chromosomes 2 to 18 (arrow) are not depicted). The fraction of the recipient strain C57BL/6 on the diploid genome is indicated for each animal and each marker. The desired homozygous genotype (2× allele of the recipient strain) is indicated by the value 1.0, and the heterozygous genotype (allele of the donor strain + allele of the recipient strain) is indicated by the value 0.5. The experimentally derived average value of the fraction of the recipient strain on the diploid genome is 76.9 % (red; theoretically expected: 75 %) for the animals N2-1 to N2-10. Animal N2-9 shows the highest fraction, the value of 84.1 % (yellow) is clearly above the mean value.

3.5 Number of backcross generations

A speed-congenic strain can be established within 5 - 6 backcross generations and will then harbour a similar rate of contamination of unlinked genome loci as conventionally bred congenic strains (Tab. 1). This is an enormous saving of time. However, computer simulations resulted in a relatively high risk of the appearance of undetected unlinked genome fragments of the donor strain in these speed-congenic strains (Armstrong et al. 2006). Thus, it is recommended to carry out 1 - 2 additional backcrosses to the recipient strain after having carried out a MASP to establish a speed-congenic strain. In addition, offspring of early backcross generations of speed-congenic strains still harbour large genomic fragments of the donor strain around the selected allele, in the case that the linked genomic fragments of the donor strain have not been included to the MASP. The mean length of this fragment is about 40 cM in the generation N5.

4. Practical problems in the generation of „speed-congenic" strains

4.1 Non-informative markers of the chosen strains

Databases are published on the internet providing information on polymorphic markers of inbred strains (see 6.). However, experimenters also use laboratory-specific substrains or inbred strains not listed. This may lead to the situation that not all chosen markers are informative in the own experiments. Testing the markers on the strains which should be used for the generation of congenic strains and on the generation N1 (= F1) bred thereof may circumvent this problem.

4.2 Unequal distribution of the informative markers among the genome

Markers should be chosen with equal distribution over the whole genome and a mean distance of ca. 10 - 25 cM (see 3.2). Databases do not deliver useful markers in every case especially for microsatellite markers. Thus, informative microsatellites and SNPs may be combined for the genotyping protocol. Alternatively, a search for additional microsatellites and SNPs can be carried out using databases of the genomic sequences of mouse and rat, and the respective primers may be designed and tested in the chosen strains of the project.

4.3 Low reproductive performance of the animals

A prerequisite for the establishment of speed-congenic strains with the use of a MASP is the appearance of a sufficient number of offspring for the genotyping and the successful mating This requires a sufficient number of breeding pairs harbouring a sufficiently high reproductive performance.

In some cases, the reproductive performance may be negatively influenced by various parameters like genetic background, mutations, environmental factors, or hygiene status. This may impair breeding performance, litter size and rearing of the offspring. The colony index1 of the strains used in the project may give first hints for potential future reproductive problems. In addition, further problems may include an aberrant ratio of the genotypes or an aberrant sex ratio in the backcross generations.

4.4 Practical examples

Examples of practical problems:

- The mutant gene $Dpp4^{m}$ was planned to be bred from the inbred donor strain F344-$Dpp4^{m}$ to the novel inbred recipient strains DA and LEW.Cg-$RT1^{av1}$. For both genetic combinations (F344-$Dpp4^{m}$ vs. DA and F344-$Dpp4^{m}$ vs. LEW.Cg-$RT1^{av1}$) a sufficient number of genetic markers was available to examine the genetic background in the consecutive backcross generations. The congenic strain DA.F344-$Dpp4^{m}$ harboured the desired homogenous genetic background in generation N6. However, the breeding (F344-$Dpp4^{m}$ × LEW.Cg-$RT1^{av1}$) × LEW.Cg-$RT1^{av1}$ resulted in such low reproductive performances that it was not possible to establish generation N3. Therefore, the congenic inbred strain DA.F344-$Dpp4^{m}$ was used as an alternative donor strain to backcross the mutation $Dpp4^{m}$ to the strain LEW.Cg-$RT1^{av1}$. This resulted in the desired homogenous genetic background of the strain LEW.Cg (F344)-$RT1^{av1}$ $Dpp4^{m}$ in generation N6.
- The project of breeding the transgene Il-18bp of the mouse inbred strain B6-Tg(Il-18bp) to the strain MRL-FasLpr failed after the establishment of generation N6, because the offspring were cannibalized by their mothers in the further generations.
- 56 informative microsatellite markers were identified and used for the generation of the congenic inbred strain MRL.Cg-Fas^{lpr} $Il\text{-}18^{tm1Aki/Ztm}$. About 40 offspring were produced for each generation, where half of them were heterozygous for the mutant interleukin-18 allele on chromosome 9. The males with the highest total genomic fractions of the recipient strain were used for the further backcrosses. The absence of heterozygous loci in the genetic background of the congenic strain was identified not before generation N13. However, the genetic end control of the novel congenic strain with the use of 27 SNPs again resulted in a heterozygous locus on chromosome 7 in two animals.

5. Conclusion

The conventional generation of a congenic strain is carried out by backcross of the allele of interest usually for 10 generations from the genetic background of a donor strain to that of a recipient strain. The selection hereby is carried out exclusively for the presence of the allele of interest. The generation of speed-congenic strains additionally includes the use of a marker assisted selection protocol (MASP) for maximizing the genomic fraction of the recipient strain in each generation, starting from generation N2. With regard to the costs and profits, usually a genome-wide marker set with an average distance of the neighbouring markers of 10 - 20 cM is used (1 cM refers to about 2 megabase pairs (Mb) of the mouse and rat genome), and 10 – 20 male carriers of the allele of interest are examined for each generation. This may lead to the generation of speed-congenic strains after 5 – 6 backcross generations, which show a similar contamination by donor strain genome fractions without linkage to the allele of interest as compared to conventionally bred congenic strains. Additional markers may be used in the subsequent backcross generations after the fixation of defined genomic fragments of the recipient strain. It is urgently advised to use further markers for the reduction of the donor strain genome fragment flanking the selected allele. The generation of the speed-congenic strain may be additionally finalized by carrying out further 1 - 2 backcross generations on the recipient strain to reduce the risk of the presence of genomic fragments of the donor strain which may remain undetected with the marker set used. Moreover, the continuous further backcross of a

congenic strain to the recipient strain principally leads to further changes of the genetic background due to the additional reduction of the linked genetic contamination. Therefore, results from analyses with animals of different backcross generations of a given congenic strain are not directly comparable. Breeding and subsequent phenotypic examination of several independent congenic strains for an allele of interest by using the same recipient inbred strain may be an additional option for to tackle the problem of potentially undetected unlinked genomic fragments of the donor strain in the congenic strain. Alternatively, genetic engineering techniques may produce organisms which harbour the identical genetic attributes as congenic strains (see 1.).

6. Internet addresses for polymorphic genetic markers

Internet addresses for polymorphic genetic markers are also collected in Benavides et al. (2020, see "Supplementary material").

6.1 Mouse

Ensembl Genome Browser (https://www.ensembl.org/index.html)

The Jackson Laboratory (https://www.jax.org/research-and-faculty/resources); including

- Mouse Genome Informatics Database (http://www.informatics.jax.org); including microsatellites with primer sequences and fragment lengths (http://www.informatics.jax.org/marker)
- Mouse Phenome Database; see "Genotypes" (https://phenome.jax.org/genotypes)
- Multiple Genome Viewer (http://www.informatics.jax.org/mgv/)
- SNP Database (http://www.informatics.jax.org/snp)

Mouse Microsatellite Data Base of Japan: SHIGEN (http://shigen.nig.ac.jp/mouse/mmdbj/)

Wellcome Sanger Institute: Mouse Genomes Project (https://www.sanger.ac.uk/science/data/mouse-genomes-project)

6.2 Rat

Ensembl Genome Browser (https://www.ensembl.org/index.html)

RGD - Rat Genome Database (http://rgd.mcw.edu)

7. Literature

Armstrong NJ, Brodnicki TC, Speed TP. 2006. Mind the gap: analysis of marker-assisted breeding strategies for inbred mouse strains. Mamm Genome 17:273-287.

Beck JA, Lloyd S, Hafezparast M, Lennon-Pierce M, Eppig JT, Festing MFW, Fisher EMC. 2000. Genealogies of mouse inbred strains. Nat Genet 24:23-25.

Behringer R. 1998. Supersonic congenics? Nat Genet 18:108.

Behringer R, Gertsenstein M, Vintersten Nagy K, Nagy A. 2014. Manipulating the mouse embryo: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Benavides F, Rülicke T, Prins J-B, Bussell J, Scavizzi F, Cinelli P, Herault Y, Wedekind D. 2020. Genetic quality assurance and genetic monitoring of laboratory mice and rats: FELASA Working Group Report. Lab Anim 54:135–148.

Berry ML, Cutler Linder C. 2007. Breeding systems: considerations, genetic fundamentals, genetic background, and strain types. *In*: Fox JG, Barthold SW, Davisson MT, Newcomer CE, Quimby FW, Smith AL (eds). The mouse in biomedical research - v.1. History, wild mice, and genetics. Academic Press, Burlington, MA, p. 53-78.

Bryda EC, Riley LK. 2008. Multiplex microsatellite marker panels for genetic monitoring of common rat strains. J Am Assoc Lab Anim Sci 47:37–41.

Canzian F. 1997. Phylogenetics of the laboratory rat Rattus norvegicus. Genome Res 7:262-267.

Cox A, Ackert-Bicknell CL, Dumont BL, Ding Y, Bell JT, Brockmann GA, Wergedal JE, Bult C, Paigen B, Flint J, Tsaih SW, Churchill GA, Broman KW. 2009. A new standard genetic map for the laboratory mouse. Genetics 182:1335-1344.

Doetschman T. 2009. Influence of genetic background on genetically engineered mouse phenotypes. Methods Mol Biol 530:423–433.

Hedrich HJ. 2000. History, strains and models. *In*: Krinke GJ (ed). The laboratory rat. Elsevier, London, p 3-16.

Kaiser J. 2020. Growth spurt for height genetics. Science 370:645.

Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, Furlotte NA, Eskin E, Nellaker C, Whitley H, Cleak J, Janowitz D, Hernandez-Pliego P, Edwards A, Belgard TG, Oliver PL, McIntyre RE, Bhomra A, Nicod J, Gan X, Yuan W, van der Weyden L, Steward CA, Bala S, Stalker J, Mott R, Durbin R, Jackson IJ, Czechanski A, Guerra-Assunçao JA, Donahue LR, Reinholdt LG, Payseur BA, Ponting CP, Birney E, Flint J, Adams DJ. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477:289-294.

Kenneth NS, Younger JM, Hughes ED, Marcotte D, Barker PA, Saunders TL, Duckett CS. 2012. An inactivating caspase 11 passenger mutation originating from the 129 murine strain in mice targeted for c-IAP1. Biochem J 443:355–359.

Landa V, Zidek V, Pravenec M. 2010. Generation of rat "supersonic" congenic/conplastic strains using superovulation and embryo transfer. Methods Mol Biol 597:267-275.

Lusis AJ, Yu J, Wang SS. 2007. The problem of passenger genes in transgenic mice. Arterioscler Thromb Vasc Biol 27:2100-2103.

Markel P, Shu P, Ebeling C, Carlson GA, Nagle DL, Smutko JS, Moore KJ. 1997. Theoretical and empirical issues for marker-assisted breeding of congenic mouse strains. Nat Genet 17:280-284.

Morgan AP, Fu C-P, Kao C-Y, Welsh CE, Didion JP, Yadgary L, Hyacinth L, Ferris MT, Bell TA, Miller DR, Giusti-Rodriguez P, Nonneman RJ, Cook KD, Whitmire JK, Gralinski LE, Keller M, Attie AD, Churchill GA, Petkov P, Sullivan PF, Brennan JR, McMillan L, Pardo-Manuel de Villena F. 2016. The mouse universal genotyping array: from substrains to subspecies. G3 (Bethesda) 6:263-279.

Ogonuki N, Inoue K, Hirose M, Miura I, Mochida K, Sato T, Mise N, Mekada K, Yoshiki A, Abe K, Kurihara H, Wakana S, Ogura A. 2009. A high-speed congenic strategy using first-wave male germ cells. PLoS ONE 4:e4943.

Petkov PP, Ding Y, Cassell MA, Zhang W, Wagner G, Sargent EE, Asquith S, Crew V,Johnson KA, Robinson P, Scott VE, Wiles MV. 2004. An efficient SNP system for mouse genome scanning and elucidating strain relationships. Genome Res 14:1806-1811.

Schalkwyk LC, Jung M, Daser A, Weiher M, Walter J, Himmelbauer H, Lehrach H. 1999. Panel of microsatellite markers for whole-genome scans and radiation hybrid mapping and a mouse family tree. Genome Res 9:878-887.

Silver LM. 1995. Mouse genetics: concepts and applications. Oxford University Press, New York, NY. http://www.informatics.jax.org; see “More resources”

STAR Consortium. 2008. SNP and haplotype mapping for genetic analysis in the rat. Nat Genet 40:560-566.

Szatkiewicz JP, Beane GL, Ding Y, Hutchins L, Pardo-Manuel de Villena F, Churchill GA. 2008. An imputed genotype resource for the laboratory mouse. Mamm Genome 19:199-208.

Taft RA, Davisson M, Wiles MV. 2006. Know thy mouse. Trends Genet 22:649-53.

Uchimura A, Higuchi M, Minakuchi Y, Ohno M, Toyoda A, Fujiyama A, Miura I, Wakana S, Nishino J, Yagi T. 2015. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. Genome Res 25:1125-1134.

Vanden Berghe T, Hulpiau P, Martens L, Vandenbroucke RE, Van Wonterghem E, Perry SW, Bruggeman I, Divert T, Choi SM, Vuylsteke M, Shestopalov VI, Libert C, Vandenabeele P. 2015. Passenger mutations confound interpretation of all genetically modified congenic mice. Immunity 43:200–209.

Visscher PM. 1999. Speed congenics: accelerated genome recovery using genetic markers. Genet Res 74:81-85.

Wakeland E, Morel L, Achey K, Yui M, Longmate J. 1997. Speed congenics: a classic technique in the fast lane (relatively speaking). Immunol Today 18:472-477.

Weil MM, Brown BW, Serachitopol DM. 1997. Genotype selection to rapidly breed congenic strains. Genetics 146:1061-1069.

Witmer PD, Doheny KF, Adams MK, Boehm CD, Dizon JS, Goldstein JL, Templeton TM, Wheaton AM, Dong PN, Pugh EW, Nussbaum RL, Hunter K, Kelmenson JA, Rowe LB, Brownstein MJ. 2003. The development of a highly informative mouse simple sequence length polymorphism (SSLP) marker set and construction of a mouse family tree using parsimony analysis. Genome Res 13:485-491.

Wong GT. 2002. Speed congenics: applications for transgenic and knock-out mouse strains. Neuropeptides 36:230-236.

Yang H, Ding Y, Hutchins LN, Szatkiewicz J, Bell TA, Paigen BJ, Graber JH, Pardo-Manuel de Villena F, Churchill GA. 2009. A customized and versatile high-density genotyping array for the mouse. Nat Methods 6:663–666.

Disclaimer

Any use of GV-SOLAS publications (specialist information, statements, booklets, recommendations, etc.) and application of the information contained therein are at the express risk of the user. Neither GV-SOLAS nor also the authors can accept liability for any accidents or damages of any kind arising from the use of a publication (e.g. resulting from the absence of safety instructions), irrespective of legal grounds. Liability claims against GV-SOLAS and the author for damages of a material or non-material nature caused by the use or non-use of the information or by the use of erroneous and/or incomplete information are in principle excluded. Legal claims and claims for damages are therefore excluded. The work, including all content, was compiled with utmost care. However, GV-SOLAS and the authors assume no responsibility and no liability for the currentness, correctness, completeness or quality of the information provided or for printing errors. GV-SOLAS and the authors accept no legal responsibility or liability in any form for incorrect statements and consequences arising therefrom. Responsibility for the content of the internet pages printed in these publications lies solely with the owner of the websites concerned. GV-SOLAS and the authors have no influence on the design and content of third-party websites and therefore distance themselves from all third-party content. Responsibility within the meaning of press legislation lies with the board of GV-SOLAS.