Report on a problem studied at the UK Mathematics-in-Medicine Study Group NC3Rs Study Group 2013
< http://www.maths-in-medicine.org/uk/2013-nc3rs/mouse-genetics/ >

Using mathematical modelling to optimise work flow in the Sanger Mouse Genetics Project

Problem presenter: Chris Lelliott
Contributors: Chris Lelliott, Shailesh Naire, Jennifer Salisbury, Marcus Tindall, Hannah Wardle-Jones, Jacqui White
Report prepared by: Chris Lelliott, Shailesh Naire, Marcus Tindall

1 Introduction

In recent years, the generation and phenotypic analysis of genetically modified mice has been increasing at breakneck pace around the world. A number of consortia operating under the banner of the International Mouse Phenotyping Consortium (IMPC - http://www.mousephenotype.org/) have the goal to produce and phenotype mice with null alleles for all known and predicted coding genes [1, 2]. The Sanger Mouse Genetics Project (MGP) is one of the world leaders in this field and operates well established large-scale and high-throughput screening (HTS) pipeline for the generation, expansion, phenotyping and export of colonies of genetically-modified mice. This entails the coordination of several steps and teams to enable the delivery of mice for downstream applications. So far the MGP has phenotyped and exported over 500 lines of mice, making it one of the largest non-commercial mouse breeding facilities in the UK. Identifying and resolving issues and weaknesses in the pipeline is a key area of work in the MGP. For example, we have already refined data generation and analysis to maximize the value of results from the animals used in our pipeline [3]. Due to the complexity and scale of the project, any inefficiency which exists in our processes will increase the number of surplus mice, extend the transit time of colonies through the pipeline and result in the suboptimal use of space and resources within the facility.

To our knowledge, mathematical modelling has not been applied to the workflows of biological *in vivo* HTS projects. The intrinsic variability of the properties of model organisms do not intuitively lend themselves to current methods of pipeline and work-flow analysis. Nonetheless, optimization of the pipeline will generate significant benefits from the 3Rs perspective, as well as improving resource usage and distribution. Given the global scale of phenotyping efforts and the increasing use of high-throughput animal screens, lessons learned from the optimisation process for the MGP will have wide-ranging benefits for animal usage and welfare around the world.

In a standard high-throughput pipeline, such as in a manufacturing plant or logistics network, we know the absolute values and variability for the transit time and capacity for each of these steps with a high degree of certainty. Linear programming and related methods can be used to identify the optimal outcome for a given set of values, such as those derived from a pipeline. In the MGP

pipeline (see Figure 1), for each line of mice with a specific mutation (known as a colony) a specific number and type of mice is required for each of the end points. In order to achieve this, each colony needs to be expanded from mice that have been proven to have undergone germ line transmission of the allele; i.e. the mutant allele is stable within the genome and can be transmitted to successive generations of progeny. The colony is only complete once the required delivery of mice to all end points has occurred and the results from the end point have been obtained and quality controlled. The current target is to complete at least 160 colonies per year in the main phenotyping pipelines.

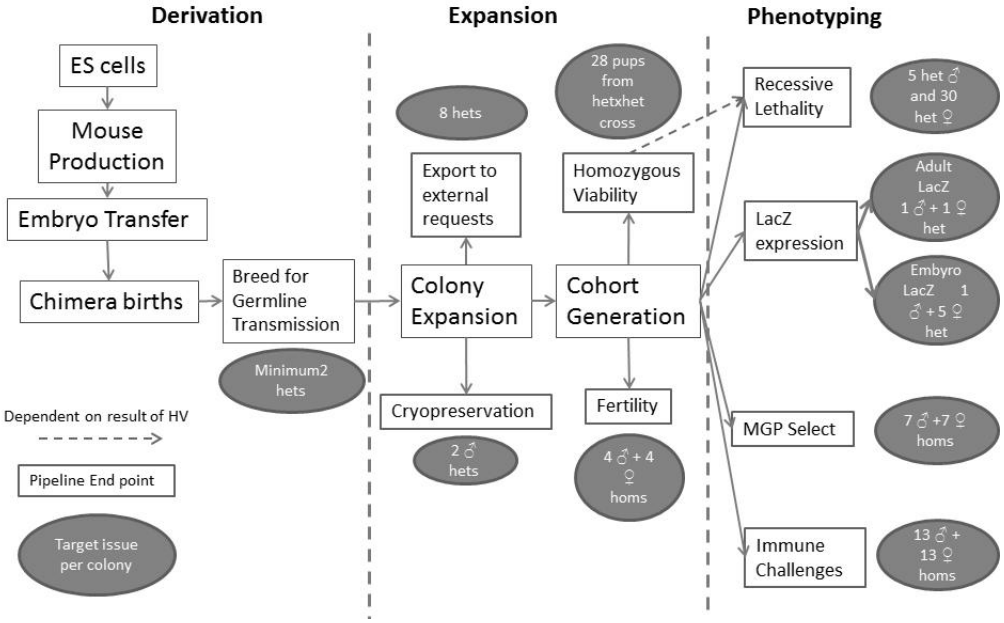

Figure 1: A schema of the current MGP pipeline. Chimaeric mice are initially derived from ES cells containing the mutant allele to be studied. Some of these chimaeras are able to transfer the mutant allele to their progeny (known as germline transmission) to establish a stable allele for the generation of the colony. The colony then undergoes a series of expansions to generate cohorts of mice for delivery to experimental (e.g. Immune challenges, MGP Select) and operational endpoints (Cryopreservation, export). Depending on requirements some colony deliveries such as export are not required. Dependencies also exist so that mice that have a sub- or non-viability phenotype also enter Recessive Lethality.

The biological nature of the pipeline means that for a given colony, each step is intrinsically variable despite specific goals for the pipeline that need to be achieved. The properties of a particular colony cannot accurately be determined prior to the entry of the colony into the pipeline, although a prediction may be made *a priori* based on knowledge of gene function and the genetic background the colony is maintained on. Colonies may display homozygous sub-fertility (around 5% of all tested) or are homozygous lethal or sub-viable (HV; approximately 40%), extending the time needed for generating mice for downstream phenotyping. There are also dependencies between endpoints. For example, mice are only delivered to Recessive Lethality if they are shown to be homozygous lethal or subviable. A variable proportion of mice are also lost at each step due to the nature of the genetic modification and to the background mortality rate of the mice. In addition, the critical parameters determining the behaviour of individuals within the colony (e.g. breeding success,

longevity, penetration of any deleterious phenotype) will also differ because of the natural variation of individuals within a population. These combined factors mean that some degree of oversupply of mice is required to meet the demand for each end point.

One solution to overcome the variability in the pipeline is to generate and fill the pipeline with as many mice and colonies as possible, despite not knowing which colonies are sub-fertile, sub-viable or prone to loss. However, it is unlikely that any facility would have the space available for such a strategy, especially in the context of a high-throughput pipeline, irrespective of the cost in terms of human resources for animal husbandry and budgetary restraints. In addition, during periods where loss is minimal and cohort generation is optimal, there is the risk of overwhelming available resources and animal wastage would be markedly increased.

In contrast, under-supply of mice so that end-points are not filled in the manner desired for statistically powered scientific analysis will result in further breeding and mouse generation to gain adequate numbers for replication of the end-point. An additional compounding factor is that mouse fertility decreases with age, which means that the longer a colony is needed for different end points, the more breeding that are required to resupply mice for cohort generation in addition to pipeline end-points. From a 3Rs perspective, there needs to be a balance between generating enough mice to complete the pipeline, whilst avoiding wastage through over- or under-supply to each endpoint. A strategy that can optimise the generation and expansion of breeding cohorts of mice to allow a consistent flow into the pipelines, which is also aligned with overall goals and resource constraints, is required.

Despite significant biological variability within each step of the pipeline, we are able to predict the genotypes of mice within the pipeline using the laws of inheritance defined by Gregor Mendel (see Figure 2). Therefore, it should be possible to mathematically model colony expansion using a range of different mating combinations. Based on this modelling, a number of suggestions for how best to utilize the mice in the colony for optimal results can be obtained.

Parent 1	Parent 2	Offspring WT	Offspring Het	Offspring Hom
WT	WT	100%	0%	0%
WT	Het	50%	50%	0%
WT	Hom	0%	100%	0%
Het	Het	25%	50%	25%
Het	Hom	0%	50%	50%
Hom	Hom	0%	0%	100%

Figure 2: Mendelian inheritance ratios for mice containing a single mutant allele located on an autosomal chromosome. WT, wild-type; HET, heterozygous (one copy of the mutant allele); HOM, homozygous (two copies of the mutant allele). Therefore a mating of a WT parent and a HET parent will produce on average 50% WT and 50% HET progeny. To obtain sex-specific genotypes, the ratio should be divided by two.

2 Mathematical model formulation

We focus on the expansion part of the MGP pipeline, i.e., the generation of wild-type (WT), homozygous (HOM) and heterozygous (HET) mice cohorts. The number of males and females of the population of each species phenotype at generation i is represented by $M_i^{WT,HET,HOM}$ and $F_i^{WT,HET,HOM}$, respectively. The change in the number of population between generations is due to:

(a) co-operative interaction between the males and females each species phenotype producing offspring. The outcome is based on Mendel's gene inheritance laws shown in Figure 2. It is assumed that the interaction between males and females is well-mixed.

(b) removal due to death, infertility, gene defect, etc.

As mentioned above, the interactions between the species phenotype is described by Mendel's inheritance laws and is schematically written as follows.

$$M_i^{WT} + F_i^{WT} \xrightarrow{r_1} a_1 M_{i+1}^{WT} + b_1 F_{i+1}^{WT}, \tag{1a}$$
$$M_i^{WT} + F_i^{HET} \xrightarrow{r_2} a_2 M_{i+1}^{WT} + b_2 F_{i+1}^{WT}, \tag{1b}$$
$$M_i^{WT} + F_i^{HET} \xrightarrow{r_5} a_5 M_{i+1}^{HET} + b_5 F_{i+1}^{HET}, \tag{1c}$$
$$M_i^{HET} + F_i^{WT} \xrightarrow{r_3} a_3 M_{i+1}^{WT} + b_3 F_{i+1}^{WT}, \tag{1d}$$
$$M_i^{HET} + F_i^{WT} \xrightarrow{r_6} a_6 M_{i+1}^{HET} + b_6 F_{i+1}^{HET}, \tag{1e}$$
$$M_i^{HET} + F_i^{HET} \xrightarrow{r_7} a_7 M_{i+1}^{HET} + b_7 F_{i+1}^{HET}, \tag{1f}$$
$$M_i^{HET} + F_i^{HET} \xrightarrow{r_4} a_4 M_{i+1}^{WT} + b_4 F_{i+1}^{WT}, \tag{1g}$$
$$M_i^{HET} + F_i^{HET} \xrightarrow{r_{12}} a_{12} M_{i+1}^{HOM} + b_{12} F_{i+1}^{HOM}, \tag{1h}$$
$$M_i^{WT} + F_i^{HOM} \xrightarrow{r_9} a_9 M_{i+1}^{HET} + b_9 F_{i+1}^{HET}, \tag{1i}$$
$$M_i^{HOM} + F_i^{WT} \xrightarrow{r_8} a_8 M_{i+1}^{HET} + b_8 F_{i+1}^{HET}, \tag{1j}$$
$$M_i^{HET} + F_i^{HOM} \xrightarrow{r_{11}} a_{11} M_{i+1}^{HET} + b_{11} F_{i+1}^{HET}, \tag{1k}$$
$$M_i^{HET} + F_i^{HOM} \xrightarrow{r_{13}} a_{13} M_{i+1}^{HOM} + b_{13} F_{i+1}^{HOM}, \tag{1l}$$
$$M_i^{HOM} + F_i^{HET} \xrightarrow{r_{10}} a_{10} M_{i+1}^{HET} + b_{10} F_{i+1}^{HET}, \tag{1m}$$
$$M_i^{HOM} + F_i^{HET} \xrightarrow{r_{14}} a_{14} M_{i+1}^{HOM} + b_{14} F_{i+1}^{HOM}, \tag{1n}$$
$$M_i^{HOM} + F_i^{HOM} \xrightarrow{r_{15}} a_{15} M_{i+1}^{HOM} + b_{15} F_{i+1}^{HOM}. \tag{1o}$$

In the above, r_j is the average number of successful matings between a male and a female of each species phenotype of a particular generation (successful in the sense of producing offspring). It is assumed to be constant over each generation. One can also interpret r_j as the probability of a successful mating between a male and a female member of each species phenotype. a_j, b_j are the average number of male and female offspring, respectively, for a given generation (assumed constant over each generation).

Equation (1a) describes the outcome of a WT-WT interaction, Eqs. (1b, 1c) describe the possible outcomes of a male WT and female HET interaction, Eqs. (1d, 1e) describe the possible outcomes

of a male HET and female WT interaction, Eqs. (1f, 1g, 1h) describe the possible outcomes of a HET-HET interaction, Eq. (1i) describes the possible outcome of a male WT and female HOM interaction, Eq. (1j) describes the possible outcome of a male HOM and female WT interaction, Eqs. (1k, 1l) describe the possible outcomes of a male HET and female HOM interaction, Eqs. (1m, 1n) describe the possible outcomes of a male HOM and female HET interaction and Eq. (1o) describes the possible outcome of a HOM-HOM interaction.

The average number of male and female offspring for a given generation, a_j, b_j, cannot be independently prescribed. They are constrained by Mendel's gene inheritance laws as follows.

$$a_1 + b_1 = N_{WT}^{WT}, \tag{2a}$$
$$a_2 + b_2 = \frac{1}{2} N_{HET}^{WT}, \tag{2b}$$
$$a_5 + b_5 = \frac{1}{2} N_{HET}^{WT}, \tag{2c}$$
$$a_3 + b_3 = \frac{1}{2} N_{WT}^{HET}, \tag{2d}$$
$$a_6 + b_6 = \frac{1}{2} N_{WT}^{HET}, \tag{2e}$$
$$a_4 + b_4 = \frac{1}{4} N_{HET}^{HET}, \tag{2f}$$
$$a_7 + b_7 = \frac{1}{2} N_{HET}^{HET}, \tag{2g}$$
$$a_{12} + b_{12} = \frac{1}{2} N_{HET}^{HET}, \tag{2h}$$
$$a_9 + b_9 = \frac{1}{2} N_{HOM}^{WT}, \tag{2i}$$
$$a_8 + b_8 = \frac{1}{2} N_{HOM}^{WT}, \tag{2j}$$
$$a_{11} + b_{11} = \frac{1}{2} N_{HOM}^{HET}, \tag{2k}$$
$$a_{13} + b_{13} = \frac{1}{2} N_{HOM}^{HET}, \tag{2l}$$
$$a_{10} + b_{10} = \frac{1}{2} N_{HET}^{HOM}, \tag{2m}$$
$$a_{14} + b_{14} = \frac{1}{2} N_{HET}^{HOM}, \tag{2n}$$
$$a_{15} + b_{15} = N_{HOM}^{HOM}, \tag{2o}$$

where N_{WT}^{WT}, N_{HET}^{WT}, N_{WT}^{HET}, N_{HET}^{HET}, N_{HOM}^{WT}, N_{HOM}^{HET}, N_{HET}^{HOM} and N_{HOM}^{HOM} are the total number of offspring due to each successful interaction for each generation and are assumed constant over each generation.

Based on the above, the change in the number of population of each species phenotype between generations is modelled as follows. The change in the number of population of the male species

phenotype is

$$\begin{aligned} M_{i+1}^{WT} - M_i^{WT} &= a_1 r_1 M_i^{WT} F_i^{WT} + a_2 r_2 M_i^{WT} F_i^{HET} + a_3 r_3 M_i^{HET} F_i^{WT} \\ &\quad + a_4 r_4 M_i^{HET} F_i^{HET} - d_1 M_i^{WT}, \qquad (3a) \\ M_{i+1}^{HET} - M_i^{HET} &= a_5 r_5 M_i^{WT} F_i^{HET} + a_6 r_6 M_i^{HET} F_i^{WT} + a_7 r_7 M_i^{HET} F_i^{HET} \\ &\quad + a_8 r_8 M_i^{HOM} F_i^{WT} + a_9 r_9 M_i^{WT} F_i^{HOM} + a_{10} r_{10} M_i^{HOM} F_i^{HET} \\ &\quad + a_{11} r_{11} M_i^{HET} F_i^{HOM} - d_2 M_i^{HET}, \qquad (3b) \\ M_{i+1}^{HOM} - M_i^{HOM} &= a_{12} r_{12} M_i^{HET} F_i^{HET} + a_{13} r_{13} M_i^{HET} F_i^{HOM} \\ &\quad + a_{14} r_{14} M_i^{HOM} F_i^{HET} + a_{15} r_{15} M_i^{HOM} F_i^{HOM} - d_3 M_i^{HOM}. \qquad (3c) \end{aligned}$$

Similarly, the change in the number of population of the female species phenotype is

$$\begin{aligned} F_{i+1}^{WT} - F_i^{WT} &= b_1 r_1 M_i^{WT} F_i^{WT} + b_2 r_2 M_i^{WT} F_i^{HET} + b_3 r_3 M_i^{HET} F_i^{WT} \\ &\quad + b_4 r_4 M_i^{HET} F_i^{HET} - d_4 F_i^{WT}, \qquad (4a) \\ F_{i+1}^{HET} - F_i^{HET} &= b_5 r_5 M_i^{WT} F_i^{HET} + b_6 r_6 M_i^{HET} F_i^{WT} + b_7 r_7 M_i^{HET} F_i^{HET} \\ &\quad + b_8 r_8 M_i^{HOM} F_i^{WT} + b_9 r_9 M_i^{WT} F_i^{HOM} + b_{10} r_{10} M_i^{HOM} F_i^{HET} \\ &\quad + b_{11} r_{11} M_i^{HET} F_i^{HOM} - d_5 F_i^{HET}, \qquad (4b) \\ F_{i+1}^{HOM} - F_i^{HOM} &= b_{12} r_{12} M_i^{HET} F_i^{HET} + b_{13} r_{13} M_i^{HET} F_i^{HOM} \\ &\quad + b_{14} r_{14} M_i^{HOM} F_i^{HET} + b_{15} r_{15} M_i^{HOM} F_i^{HOM} - d_6 F_i^{HOM}. \qquad (4c) \end{aligned}$$

Here, d_j is a *per capita removal rate* (assumed constant).

Given an initial number of population at generation zero ($i = 0$), we can then solve the finite-difference equations (Eqs. (3,4)) to obtain the number of population of each species phenotype at the next generation.

3 Results

The governing difference equations (Eqs. (3,4)) were coded and solved in Matlab. We assume an initial population size of 10 WT (5 male and 5 female) and 4 HET (2 male and 2 female). This was based on initial estimates to the Sanger mouse production process. It was further assumed that mice in the pipeline had an 80% chance of successfully mating, regardless of phenotype (wild-type, heterozygous, homozygous), and 30% of mice were lost at each generation. This was used to estimate the constants r_j and d_j. Breeding between wild-types was ignored - this just increases the pool of wild-types which becomes so large that it is effectively constant. Simulations for only 2 generations of mating were considered; this being the average number of matings between individuals. Finally the total offspring size (N_j^i in Eq. (2)) was assumed to be 12 for offspring produced from heterozygous-heterozygous matings and 10 for the others, which was broken down into numbers of males and females (a_j and b_j in Eq. (2)).

In order to understand the contribution of each mouse phenotype (wild-type, heterozygous, homozygous) to the number of each phenotype produced in the pipeline, we initially began by breaking down the pipeline to include only the production of wild-type mice. This was then extended to

understand the effect of heterozygous mating and finally homozygous mating on the total number of each phenotype. At this initial stage the effect of any feedbacks, e.g. homozygous mating with wild-types was ignored.

Figure 3 demonstrates the increase in the number of wild-types after 2 generations of mating with heterozygous mice. The parameter values are: $r_2 = r_3 = 0.8$, whilst all other re-productive rates were set to zero, $N_{HET}^{WT} = N_{WT}^{HET} = 10$, $a_2 = a_3 = 2$, $b_2 = b_3 = 3$, $M_0^{WT} = F_0^{WT} = 5$ and $M_0^{HET} = F_0^{HET} = 2$. The number of wild-types (both male and female) rises rapidly with the number of females slightly larger than males due to the larger number of female compared to male offspring.

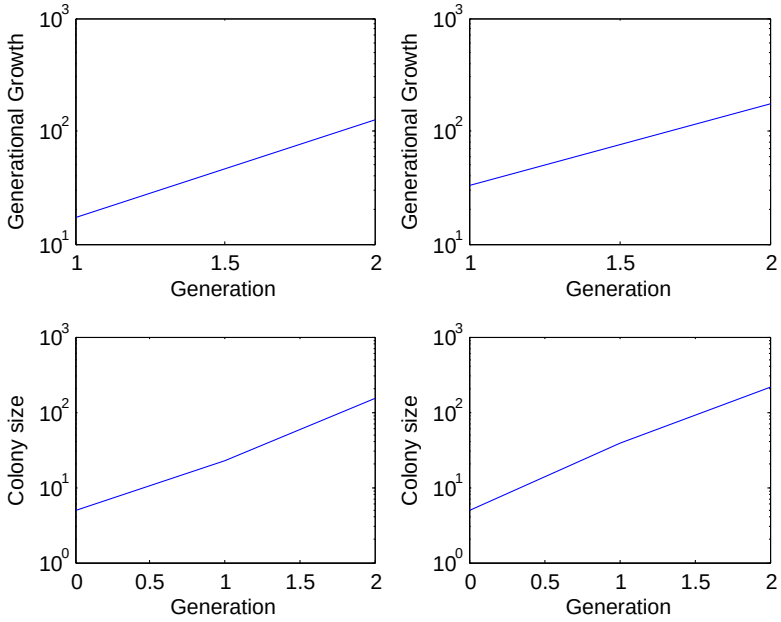

Figure 3: **The increase in wild-type mice numbers (males on the left and females on the right) as a result of breeding between wild-type and hetrozygous mice**. The parameter values are: $r_2 = r_3 = 0.8$, whilst all other re-productive rates were set to zero, $N_{HET}^{WT} = N_{WT}^{HET} = 10$, $a_2 = a_3 = 2$, $b_2 = b_3 = 3$, $M_0^{WT} = F_0^{WT} = 5$ and $M_0^{HET} = F_0^{HET} = 2$.

The next step was to understand how hetrozygous mating with wild-types affected the number of wild-types and heterozygous which is shown in Figure 4. The parameter values are: $r_2 = r_3 = r_5 = r_6 = 0.8$, whilst all other re-productive rates were set to zero, $N_{HET}^{WT} = N_{WT}^{HET} = 10$, $a_2 = a_3 = a_5 = a_6 = 2$, $b_2 = b_3 = b_5 = b_6 = 3$, $M_0^{WT} = F_0^{WT} = 5$ and $M_0^{HET} = F_0^{HET} = 2$. As can be seen in Figure 4, this breeding has a dramatic effect on the number of wild-types and heterozygous being produced, in comparison to Figure 3.

The final stage of the production pipeline is the formation of homozygous mice. We allow this to happen as a result of mating between wild-type and heterozygous and, heterozygous and heterozygous. The parameter values are: $r_2 = r_3 = r_5 = r_6 = r_{12} = 0.8$, whilst all other re-productive rates were set to zero, $N_{HET}^{WT} = N_{WT}^{HET} = 10$, $N_{HET}^{HET} = 12$, $a_2 = a_3 = a_5 = a_6 = 2$, $a_{12} = 1$, $b_2 = b_3 = b_5 = b_6 = 3$, $b_{12} = 2$, $M_0^{WT} = F_0^{WT} = 5$, $M_0^{HET} = F_0^{HET} = 2$ and $M_0^{HOM} = F_0^{HOM} = 0$. This leads to the numbers shown in Figure 5.

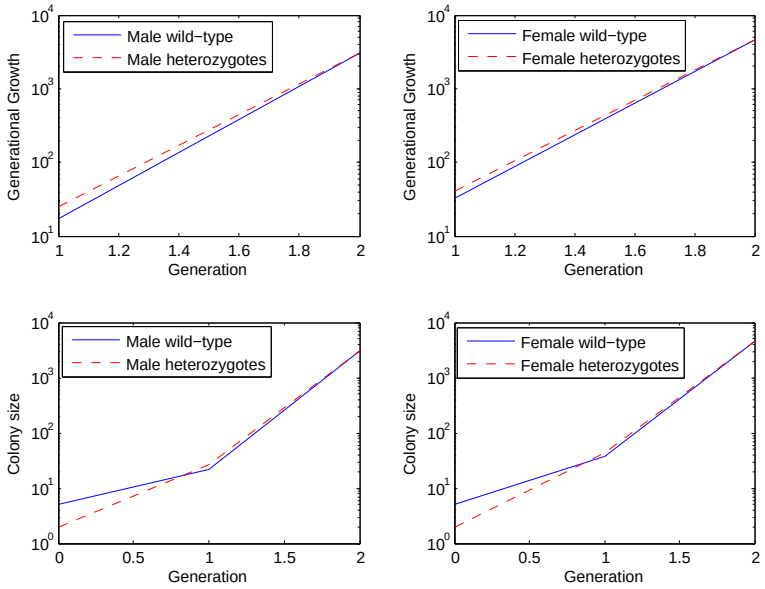

Figure 4: **The effect of including breeding between heterozygous on the wild-type and heterozygous mice populations**. In this case $r_2 = r_3 = r_5 = r_6 = 0.8$, whilst all other re-productive rates were set to zero, $N_{HET}^{WT} = N_{WT}^{HET} = 10$, $a_2 = a_3 = a_5 = a_6 = 2$, $b_2 = b_3 = b_5 = b_6 = 3$, $M_0^{WT} = F_0^{WT} = 5$ and $M_0^{HET} = F_0^{HET} = 2$.

Having provided sufficient numbers of homozygous, we then considered the effect of breeding between homozygous on their overall population numbers. The parameter values are: $r_2 = r_3 = r_5 = r_6 = r_{12} = r_{15} = 0.8$, whilst all other re-productive rates were set to zero, $N_{HET}^{WT} = N_{WT}^{HET} = N_{HOM}^{HOM} = 10$, $N_{HET}^{HET} = 12$, $a_2 = a_3 = a_5 = a_6 = 2$, $a_{12} = 1$, $a_{15} = 5$, $b_2 = b_3 = b_5 = b_6 = 3$, $b_{12} = 2$, $b_{15} = 5$, $M_0^{WT} = F_0^{WT} = 5$, $M_0^{HET} = F_0^{HET} = 2$ and $M_0^{HOM} = F_0^{HOM} = 0$. As can be seen in Figure 6, because the numbers of homozygous are effectively smaller than that of wild-type and heterozygous mice, the numbers of homozygous are only increased by a relatively small amount in comparison to those provided by the other interactions. In this way homozyous-homozygous breeding increases the number of homozygous more slowly than that of other interactions and thus acts as a "fine-tuner" on homozygous numbers.

The final stage of our model analysis was to consider what effect breeding between the heterozygous and homozygous populations had on the total number of homozygous. The parameter values are: $r_2 = r_3 = r_5 = r_6 = r_{12} = r_{13} = r_{14} = r_{15} = 0.8$, whilst all other re-productive rates were set to zero, $N_{HET}^{WT} = N_{WT}^{HET} = N_{HOM}^{HOM} = N_{HOM}^{HET} = N_{HET}^{HOM} = 10$, $N_{HET}^{HET} = 12$, $a_2 = a_3 = a_5 = a_6 = a_{13} = a_{14} = 2$, $a_{12} = 1$, $a_{15} = 5$, $b_2 = b_3 = b_5 = b_6 = b_{13} = b_{14} = 3$, $b_{12} = 2$, $b_{15} = 5$, $M_0^{WT} = F_0^{WT} = 5$, $M_0^{HET} = F_0^{HET} = 2$ and $M_0^{HOM} = F_0^{HOM} = 0$. As can be seen in Figure 7 this feedback interaction leads to a larger increase in the homozygous population than just the breeding between homozygous. This is because there are more heterozygous available for the homozygous to mate with and the likelihood, and this interaction, via the Mendelian table, increases the number of homozygous formed.

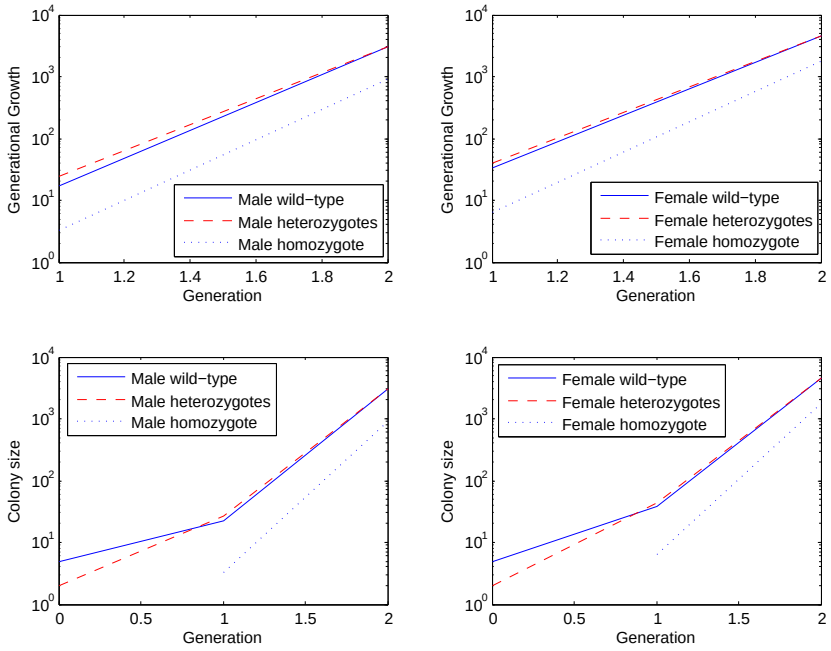

Figure 5: **The increase in homozygous numbers as a result of breeding between wild-types and heterozygous, and heterozygous and heterozygous**. The parameter values are: $r_2 = r_3 = r_5 = r_6 = r_{12} = 0.8$, whilst all other re-productive rates were set to zero, $N_{HET}^{WT} = N_{WT}^{HET} = 10$, $N_{HET}^{HET} = 12$, $a_2 = a_3 = a_5 = a_6 = 2$, $a_{12} = 1$, $b_2 = b_3 = b_5 = b_6 = 3$, $b_{12} = 2$, $M_0^{WT} = F_0^{WT} = 5$, $M_0^{HET} = F_0^{HET} = 2$ and $M_0^{HOM} = F_0^{HOM} = 0$.

4 Conclusions and future work

One of the key aims of our work has been to understand and help optimise the number of mice delivered, in respect of phenotype, to different parts of the Sanger mouse production pipeline. In this report we have taken a population based approach formulated using nonlinear difference equations to tackle this problem.

Our work has demonstrated the importance of exploiting Mendelian genetics in respect of maximising the number of both heterozygous and homozygous mice. In particular we have shown that once a number of heterozygous and homozygous mice have been produced, large increases in the number of homozygous can be obtained by breeding homozygous with heterozygous. In contrast, breeding between homozygous slowly increases the homozygous population numbers. The results of our work advocate two strategies for optimising homozygous mouse numbers. If a large number of homozygous is required, then breeding between homozygous and heterozygous should be used. If small numbers or "fine-tuning" of a large population of homozygous is required, breeding between homozygous should be utilised.

Our approach here has been on the population scale and has not accounted for small numbers of mice being bred for subsets of the heterozygous and homozygous populations. Likewise the affect of breeding between litters and use of new and old breeders has not been addressed. Such points need to be addressed in future work and other mathematical modelling methodologies, such as stochastic operational research methods, considered.

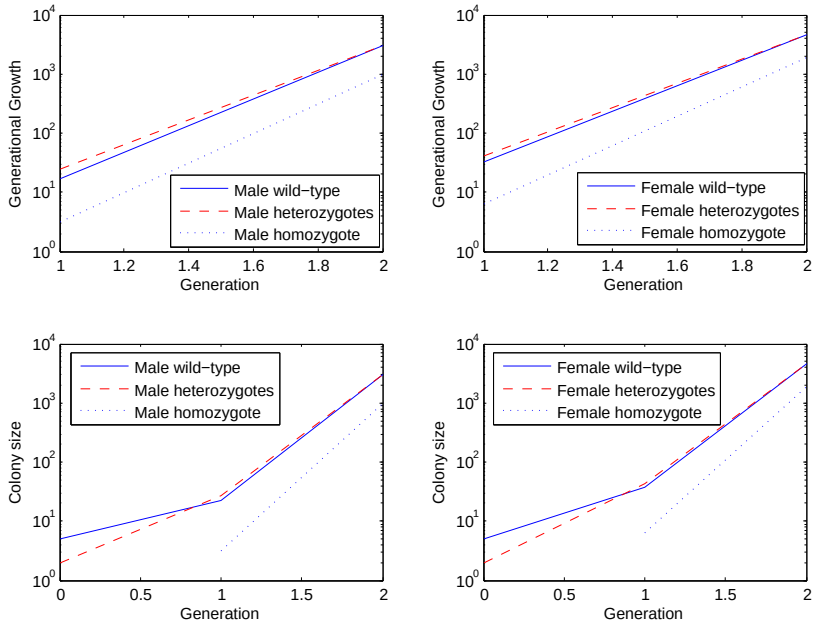

Figure 6: **Slowly increasing homozygous numbers by including breeding between homozygous**. This breeding acts as a "fine-tuner" on the total number of homozygous in the production pipeline. The parameter values are: $r_2 = r_3 = r_5 = r_6 = r_{12} = r_{15} = 0.8$, whilst all other re-productive rates were set to zero, $N_{HET}^{WT} = N_{WT}^{HET} = N_{HOM}^{HOM} = 10$, $N_{HET}^{HET} = 12$, $a_2 = a_3 = a_5 = a_6 = 2$, $a_{12} = 1$, $a_{15} = 5$, $b_2 = b_3 = b_5 = b_6 = b_{13} = b_{14} = 3$, $b_{12} = 2$, $b_{15} = 5$, $M_0^{WT} = F_0^{WT} = 5$, $M_0^{HET} = F_0^{HET} = 2$ and $M_0^{HOM} = F_0^{HOM} = 0$.

5 3Rs Impact

The Wellcome Trust Sanger MGP is a large-scale and high throughput biological pipeline, generating and phenotyping 200 colonies of genetically-modified (GM) mice per week with new entries into the pipeline at 70 mice per week. The main focus is to reduce wastage of GM mice during the breeding and expansion stage of the pipeline. This wastage arises due to any inefficiency in the process resulting in the generation of surplus mice unnecessary to meet a pipeline endpoint. These mice also consume storage space and available resources which could be used for productive breeding events. Due to demands of such high throughput phenotyping, it is believed that even a 1% optimization during this stage would significantly reduce wastage of mice in the order of several hundred per year at Sanger MGP. Moreover, this reduction would be even more significant when considering that the stated aim of the IMPC is to produce and phenotype one line for each known mouse genes, currently estimated to be 20,000, by 2021. This when considered over the seventeen research institutions which form the IMPC, would then be in the order of several thousand. Furthermore, given the ongoing global scale of phenotyping efforts in many animal laboratories world-wide, this optimization is forecast to have a huge impact in reducing mice wastage in the order of thousands.

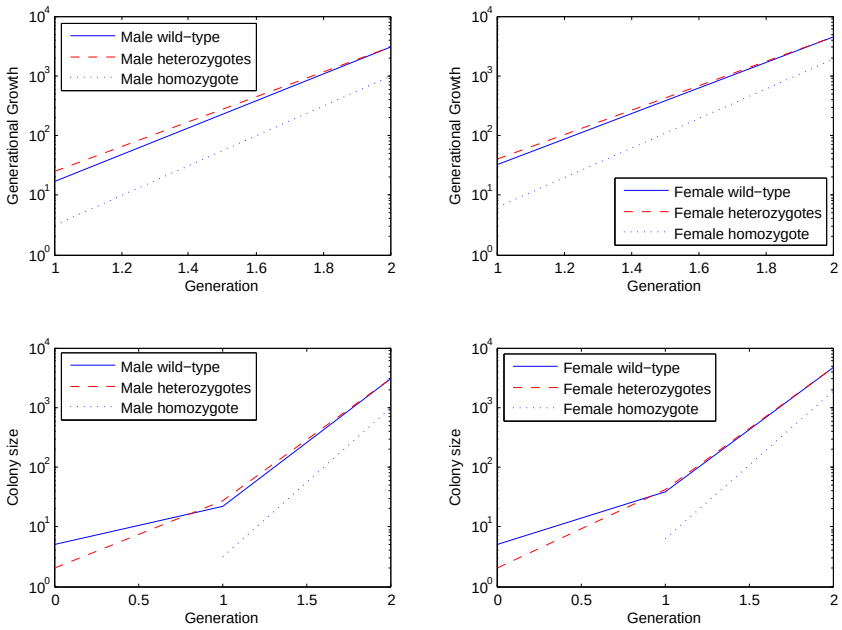

Figure 7: **The effect of breeding between homozygous and heterozygous on increasing the homozygous population numbers**. This breeding is able to increase the homozygous population numbers more rapidly than that of breeding between homozygous because of higher numbers of heterozygous and the likelihood of homozygous being born as a result of Mendelian genetics. The parameter values are: $r_2 = r_3 = r_5 = r_6 = r_{12} = r_{13} = r_{14} = r_{15} = 0.8$, whilst all other re-productive rates were set to zero, $N_{HET}^{WT} = N_{WT}^{HET} = N_{HOM}^{HOM} = N_{HOM}^{HET} = N_{HET}^{HOM} = 10$, $N_{HET}^{HET} = 12$, $a_2 = a_3 = a_5 = a_6 = a_{13} = a_{14} = 2$, $a_{12} = 1$, $a_{15} = 5$, $b_2 = b_3 = b_5 = b_6 = b_{13} = b_{14} = 3$, $b_{12} = 2$, $b_{15} = 5$, $M_0^{WT} = F_0^{WT} = 5$, $M_0^{HET} = F_0^{HET} = 2$ and $M_0^{HOM} = F_0^{HOM} = 0$.

The strategy currently followed by the Sanger Institute was largely based on intuition (following the Mendelian laws of inheritance), prior experience and some trial-and-error. They were interested in (a) scientifically validating this intuitive approach through mathematical modelling and (b) optimizing this approach so as to control and fine tune the number of mice based on requirements at the end points.

Our preliminary mathematical modelling successfully addresses both points. We were able to show the key interactions which would result in generation of an initial pool of heterozygous and homozygous mice. Moreover, we also showed that one could then increase homozygous numbers significantly by mating homozygous mice with heterozygous mice. Furthermore, these homozygous numbers could be controlled and fine tuned by mating between homozygous mice only. Hence, our modelling approach is capable of controlling the number of mice generated. One can then tune the parameters in the model to reduce any over or undersupply to the endpoints reducing any wastage of mice and also control the number of homozygous mice depending on the requirements of the endpoints, hence enabling optimization of the process. Further advancement of this model by including the stochastic element will provide much better control further enhancing the optimization and reduction of mice numbers and hence having a great impact from a reduction of animal use perspective of the 3Rs.

References

[1] Ayadi, A. *et al*, 2012, Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project, Mamm Genome, 23:600-610.

[2] Ramirez-Solis, R. *et al*, 2012, Large-scale mouse knockouts and phenotypes, Wiley Interdisciplinary Rev. Syst. Biol. Med., 4:457-463.

[3] Karp, N. *et al*, 2012, Robust and sensitivity analysis of mouse knockout phenotypes, PLoS One, 12:e52410.