Within-generation mutation variance for litter size in inbred mice

Joaquim Casellas¹ and Juan F. Medrano²

¹ Genètica i Millora Animal, IRTA-Lleida, 25198 Lleida, Spain

² Department of Animal Science, University of California-Davis, Davis, CA 95616-8521, USA
Running head: Mutational variability in inbred mice

Key words: genetic variability, genetic stability, inbred strain, mutation, C57BL/6J

Corresponding author:

Juan F. Medrano
Department of Animal Science
University of California-Davis
One Shields Ave.
Davis, CA 95616-8521, USA

Tel. (530) 752-6786
Fax. (530) 752-0175
E-mail: jfmedrano@ucdavis.edu
ABSTRACT

The mutational input of genetic variance per generation ($\sigma^2_m$) is the lower limit of the genetic variability in inbred strains of mice, although greater values could be expected due to the accumulation of new mutations in successive generations. A mixed model analysis using Bayesian methods was applied to estimate $\sigma^2_m$ and the across-generation accumulated genetic variability on litter size in 46 generations of a C57BL/6J inbred strain. This allowed for a separate inference on $\sigma^2_m$ and on the additive genetic variance in the base population ($\sigma^2_a$). The additive genetic variance in the base generation was 0.151 and quickly decreased to almost null estimates in generation 10. On the other hand, $\sigma^2_m$ was moderate (0.035) and the within-generation mutational variance increased up to generation 14, then oscillating between 0.102 and 0.234 in remaining generations. This pattern suggested the existence of a continuous uploading of genetic variability for litter size ($h^2 = 0.045$). Relevant genetic drift was not detected in this population. In conclusion, our approach allowed for separate estimation of $\sigma^2_a$ and $\sigma^2_m$ within the mixed model framework, and the heritability obtained highlighted the significant and continuous influence of new genetic variability impacting the genetic stability of inbred strains.
The importance of new mutations on polygenic variability has been suggested by several investigators in the last decades (Hill, 1982a,b; Caballero et al., 1991; Keightley, 1998). Direct evidence of new mutations with large effects in experimental selection lines were initially reported during the second half of the twentieth century (MacArthur 1949; Yoo 1980; Bradford and Famula 1984). The mutational input of genetic variance per generation ($\sigma_m^2$) can be viewed as the ultimate source of polygenic variation and thus, as the raw material for the maintenance of genetic variability in populations (Hill 1982a). Estimates of mutational heritabilities ($h_m^2 = \sigma_m^2 / (\sigma_m^2 + \sigma_e^2)$, $\sigma_e^2$ being the residual variance of the trait) in animals and cereal crops, have ranged between $10^{-4}$ and $5 \times 10^{-2}$ (Lynch 1988; Houle et al. 1996). Within this context, it is well known that spontaneous mutation continually contributes new alleles to the pool of genetic variation, allowing for response to long-term artificial selection experiments in both animals (Caballero et al. 1991; Keightley 1998) and plants (Hill 2007).

Most estimates of $\sigma_m^2$ and $h_m^2$ come from experiments focused on the rate of divergence between sub-lines from a highly inbred base population (Festing 1973; Mackay et al. 1994). These studies typically assumed a stringent scenario under mutation-drift equilibrium, which does not necessarily hold in experimental populations (Hill, 1982a). Moreover, a long time is needed for generating the strains, and the analysis using the response to selection typically ignores information on covariances between relatives within lines, a proportion of which can be genetic (Keightley and Hill 1992). Alternatively, Wray (1990) developed a straightforward approach to account for mutation effects in mixed models using the numerator relationship matrix, allowing for estimation of $\sigma_m^2$ in unselected populations. This methodology has not been widely applied, although some $\sigma_m^2$ estimates have been obtained in mice (Keightley and Hill...
A topic of interest in studies with laboratory mice is the genetic homogeneity of inbred strains across generations (Taft et al. 2006; Stevens et al. 2007) and Wray’s (1990) approach seems to be the only available methodology for testing mutation effects, given that selection is avoided in these genetically controlled strains. Nevertheless, there are no published results on \( \sigma_m^2 \) in highly inbred unselected mice, and the magnitude or impact of \( \sigma_m^2 \) on the phenotypic variance remains unknown.

Taking the infinitesimal model (Fisher 1918) as the starting point, the additive genetic variance for a given phenotypic trait in a population characterizes the amount of genetic variability and, the potential change due to natural or artificial selection, or genetic drift. Interestingly, analyses performed on the “new” genetic variance originated by mutation are commonly focused on the increment of genetic variation per generation (Caballero et al. 1991; Keightley 1998), but this does not estimate the accumulated genetic variability existing between individuals. Even in highly inbred strains, the genetic variance in a given generation of interest could be viewed as the balance of an equilibrium between \( \sigma_m^2 \) coming from current and previous generations and the loss of genetic variability due to selection, genetic drift and/or inbreeding (Hill 1982a,b). Analyses of the mutation phenomenon in laboratory species have been mainly focused on \( \sigma_m^2 \) (Festing 1973; Keightley and Hill 1992; Keightley 1998), whereas the magnitude of the overall within-generation genetic variability due to the accumulation of new mutations remains unclear.

In this study, we report estimates of mutation variance on litter size in C57BL/6J mice reared for 46 generations without selection. Wray’s (1990) algorithm was modified to estimate the amount of genetic variance in the inbred base population and the increment of the per generation variance due to mutation. In addition, we applied the Bayesian approach proposed by
SORENSEN et al. (2001) to estimate the within-generation genetic variability, in order to examine if new mutation variance compensates for losses in genetic variability.

MATERIALS AND METHODS

Mice Data Source

Mouse strain and breeding scheme: A C57BL/6J inbred strain was kept in our vivarium at the University of California-Davis for 46 non-overlapping generations (G1 to G46), between October 1988 and May 2005. This strain was founded with two C57BL/6J males and six C57BL/6J females from The Jackson Laboratory (Bar Harbor, ME). Two to five generations per year were produced. Each generation involved between two and 28 males and between six and 49 females, producing an average of 21.6 litters (Table 1). In August 1995, a sub-line was derived from G21 and was maintained for five non-overlapping generations (G22b to G26b), with a large number of litters per generation (Table 1). For each generation, males and females were selected at random from the offspring of a few litters of the previous generation, and full-sibs matings were favored. Only single (one male/one female) and group matings (one male/several females) were used to avoid multiple paternities. Each male and female produced an average of 3.5 and 1.8 litters, respectively, ranging from 1 to 16 litters for males and between 1 to 5 litters for females. Note that this strain was maintained to provide stock research mice in our colony and therefore, a variable number of litters per generation was generated depending on mice demand. All mice were fed with Purina 5008 diet (Ralston Purina Co., St. Louis, MO; 23.5% protein, 6.5% fat, 3.3 Kcal/g) and water was offered ad libitum. Mice were housed in polycarbonate cages under controlled conditions of temperature (21°C ± 2°C), humidity (40–70%) and lighting (14 h light, 10 h dark, lights on at 7 AM), and managed according to the
guidelines of the American Association for Accreditation of Laboratory Animal Care (AAALAC; http://www.aaalac.org).

**Data set and trait analyzed:** Reproductive data were recorded accurately in all generations. Sire, dam, date of mating, date of birth and number of pups at birth (alive and dead) were recorded for each litter, and pups were individually marked by ear-notching at weaning (3 weeks after birth). Records were available on 1,986 litters providing 15,044 pups. This study was focused on litter size (LS) defined as the sum of live and dead pups at birth. Phenotypic records of LS ranged between one and 14 pups, with 7.58 pups per litter on average. The pedigree file included 572 males and 1,116 females with a complete knowledge of all parental relationships.

**Bayesian Analysis**

**Model:** Litter size in the C57BL/J6 strain was analyzed with the following linear mixed model,

\[ y = X\beta + Z_1 p_1 + Z_2 p_2 + Z_3 a + Z_3 m + e \]

where \( y \) was the vector of phenotypic data and \( e \) was the vector of residuals after accounting for systematic (\( \beta \)), environmental (\( p_1 \) and \( p_2 \)) and additive genetic effects (\( a \) and \( m \)). Note that \( X, Z_1, Z_2 \) and \( Z_3 \) are appropriate incidence matrices. More specifically, \( \beta \) corrected for two systematic effects, parity number of the dam with the two levels proposed by KIRKPATRICK *et al.* (1988; first parity and following parities), and generation number with 50 levels, accounting for environmental variability between generations (see FALCONER 1960). Two environmental sources of variation common to all pups were fitted to model, the effect of the sire (\( p_1 \); SCHILLING *et al.*, 1968) and the non-genetic effect of the dam (\( p_2 \)), with 572 and 1,116 levels,
respectively. Following in part W RAY (1990), the infinitesimal genetic effect \( u \); FISHER 1918) was partitioned into two terms, \( u = a + m \), the breeding value inherited from the genetic variability in the base generation (\( a \); HENDERSON 1973) and from the additional genetic variability originated by mutation (\( m \)).

**Prior distributions:** Following a standard Bayesian development, the joint posterior distribution of the mixed model outlined above was constructed by multiplying the Bayesian likelihood with the prior distribution of all parameters in the model,

\[
p(\beta, p_1, p_2, a, m, \sigma^2_{p_1}, \sigma^2_{p_2}, \sigma^2_a, \sigma^2_m, \sigma^2_e | y) \propto p(y | \beta, p_1, p_2, a, m, \sigma^2_e) p(\beta) p(p_1 | \sigma^2_{p_1}) p(p_2 | \sigma^2_{p_2}) \\
\times p(a | A, \sigma^2_a) p(m | M, \sigma^2_m) p(\sigma^2_e)
\]

where \( \sigma^2_{p_1}, \sigma^2_{p_2}, \sigma^2_a, \sigma^2_m \) and \( \sigma^2_e \) were the appropriate variance components for \( p_1, p_2, a, m \) and \( e \), respectively. \( A \) was the standard numerator relationship matrix (WRIGHT 1922) and \( M \) was the W RAY’s (1990) numerator relationship matrix adapted to accommodate the occurrence of mutations in the genome. Note that \( M \) was defined as \( \sum_{k=0}^{t} A_k \), where \( t \) is the number of generations, \( A_k \) is the numerator relationship matrix of additive genetic effects attributed to mutations arising in time unit \( k \), and \( A_0 = A \) (see Appendix). The elements of \( A_k \) are the additive genetic relationships if ancestors born in time unit \( k - 1 \) are ignored (W RAY 1990).

Litter size data was assumed to be normally distributed as follows,

\[
p(y | \beta, p_1, p_2, a, m, \sigma^2_e) \sim \mathcal{N}(X\beta + Z_1 p_1 + Z_2 p_2 + Z_3 a + Z_4 m, I_e, \sigma^2_e)
\]

with \( I_e \) being an identity matrix with dimensions equal to the number of records in \( y \). Model parameters \( \beta, p_1, p_2, a, m \) and \( \sigma^2_e \) were assumed mutually independent. A priori distributions for \( p_1 \) and \( p_2 \) were defined as multivariate normal,
\[ p(p_1 | \sigma^2_{p_1}) \sim N(0, I_{p_1} \sigma^2_{p_1}) \]
\[ p(p_2 | \sigma^2_{p_2}) \sim N(0, I_{p_2} \sigma^2_{p_2}) \]

where \( I_{p_1} \) and \( I_{p_2} \) were identity matrices with dimensions equal to the number of elements in \( p_1 \) and \( p_2 \), respectively. Invoking the infinitesimal model \((\text{Fisher} 1918)\), \( a \) and \( m \) were assumed to follow the multivariate normal distributions
\[ p(a | \sigma^2_a) \sim N(0, A \sigma^2_a) \]
\[ p(m | \sigma^2_m) \sim N(0, M \sigma^2_m) \]

Note that mutational effects are assumed a priori independent of \( a \) \((\text{Wray} 1990)\) and therefore, genetic correlation between \( a \) and \( m \) was arbitrarily fixed to 0. Improper uniform prior distributions were assumed for \( \beta , \sigma^2_{p_1} , \sigma^2_{p_2} \) and \( \sigma^2_e \) to approximate vague prior knowledge about systematic, environmental and residual sources of variation.

According to \(\text{Henderson}(1973)\), \(\text{Gianola et al.}(1989)\) and \(\text{Im et al.}(1989)\), \( \sigma^2_a \) measures additive genetic variance at linkage equilibrium in the base population \((G1; \text{see Table 1})\). Although \( \sigma^2_a \) should be null or very small in an inbred strain, if it exists, \( \sigma^2_a \) must originate from short-term mutations arising in previous generations and is highly related to \( \sigma^2_m \). It seems reasonable to expect a similar behavior for \( \sigma^2_a \) and \( \sigma^2_m \) and therefore, the same prior was assumed for both variance components. In order to evaluate the effects of a priori information on \( \sigma^2_a \) and \( \sigma^2_m \), four different scaled inverted \( \chi^2 \) prior distribution with hyperparameters \( \nu \) and \( S^2 \) were assumed and tested independently on our data set. The first prior \((\text{PR1})\) generalized the scaled \( \chi^2 \) distribution to an improper uniform distribution by setting \( \nu = -2 \) and \( S^2 = 0 \).
(Figure 1). This prior ignores previous knowledge on $\sigma_a^2$ and $\sigma_m^2$, this being a typical assumption for variance components, where the variance is allowed to take any value between 0 and the phenotypic variance. Three more priors (Table 2) were defined on the basis of information from the literature and they varied on a trial and error basis until the desired shape of the distribution was obtained (Figure 1), following in part BLASCO et al. (1998). Given the range of mutational heritabilities reviewed by LYNCH (1988; $10^{-4}$ to $5 \times 10^{-2}$) and the moderate phenotypic variance observed in our data set (4.1 pups$^2$), it seems reasonable to expect a $\sigma_m^2$ between $4.1 \times 10^{-4}$ and $2.1 \times 10^{-1}$, without disallowing for more extreme values. The second $\chi^2$ prior (PR2) illustrated a strong a priori opinion (sharp-contour distribution) about the probable distribution of the variance components, its mode being placed at the lower bound of LYNCH’s (1988) range. Prior 3 (PR3) was an attempt to cover the range of most plausible values, although a left-skewed prior (PR4) gave a vague a priori opinion of the distribution (Figure 1) of the variance components, its mode being $2.0 \times 10^{-1}$. Note that PR2 and PR4 were proper priors although they did not have a well-defined variance. See Table 2 for a detailed description of hyperparameters for all the scaled inverted $\chi^2$ priors.

Additionally, a mixed model without $a$ and $m$ effects was analyzed (PR0), with its Bayesian likelihood reduced to

$$ p(y|\beta, p_1, p_2, \sigma_e^2; \text{~N}(X\beta + Z_1p_1 + Z_2p_2, I, \sigma_e^2) ) $$

and the prior distributions for $\beta$, $p_1$, $p_2$, $\sigma_{p_1}^2$, $\sigma_{p_2}^2$ and $\sigma_e^2$ were the same as in the full model. It can be also viewed as the general model with $p(\sigma_a^2 = 0) = 1$ and $p(\sigma_m^2 = 0) = 1$, thereby allowing for testing of the biological relevance of $a$ and $m$ effects in terms of model adequacy.

**Markov chain Monte Carlo sampling:** Within a Bayesian context, inferences are made on the joint posterior distribution or, for a given parameter of interest, on the relevant marginal
posterior distribution. Given the multidimensional form of these posterior distributions, direct integration cannot be applied. Markov chain Monte Carlo (MCMC) techniques easily bypass this limitation and allow obtaining draws from the appropriate marginal posterior distribution. For the mixed model described above, samples from the marginal posterior distribution of all unknowns in the model were obtained by Gibbs sampling (Gilks et al. 1996), following the procedures described by Sorensen et al. (1994).

For each prior distribution of $\sigma_a^2$ and $\sigma_m^2$ (PR1 to PR4), as well as for PR0, three independent MCMC chains were launched, with 500,000 iterations after discarding the first 100,000 as burn-in. Convergence was confirmed on variance components by visual inspection and by the Raftery and Lewis’ (1992) approach. In order to arrive at the most preferable model, the deviance information criterion (DIC; Spiegelhalter et al. 2002) was calculated.

**Genetic Drift and Within-Generation Additive Genetic Variances**

As mentioned above, breeding mice were randomly picked from the previous generation and no selection was applied along the 46 generations. Nevertheless, the small number of mice contributing to the next generation (see Table 1) could produce genetic drift on litter size if $\sigma_a^2$ and/or $\sigma_m^2$ were not null. Within this context, changes on the within-generation average breeding value ($a$ and $m$) and environmental effects ($p_1$ and $p_2$) were evaluated with the Bayesian approach described by Sorensen et al. (1994).

Following Sorensen et al. (2001), both additive genetic and mutation variance components were estimated within generations, using data from all individuals in the population. By definition, the additive genetic value ($a_i$) of an individual randomly picked from generation $t$ is a random variable with variance as defined by Sorensen et al. (2001),
\[ \sigma_{a(t)}^2 = E(a_t^2) - [E(a_t)]^2 = \frac{1}{n_t} \sum_{i=1}^{n_t} a_{i(t)}^2 - \left( E(a_t) \right)^2, \]

where \( \sigma_{a(t)} \) is the mathematical expectation of additive genetic values in generation \( t \), \( a_{i(t)} \) is the \( i \)th additive genetic value in generation \( t \), and \( n_t \) is the number of individuals in generation \( t \).

Inferences on the within-generation additive genetic variance were made on their marginal posterior distribution, estimated via Markov chain Monte Carlo methods. A Gibbs sampler was applied following SORENSEN et al. (2001). The same approach was applied for the within-generation mutation variance.

RESULTS

Litter size in this C57BL/6J strain averaged 7.58 ± 0.05 pups per litter with substantial variability between generations, from 9.42 ± 0.29 pups (G1) to 5.78 ± 0.40 pups (G31). Larger litter sizes were observed in earlier (G1, G3 G5) and later generations (G42 to G45) with smaller values in the intermediate ones, although a relevant phenotypic trend was not observed (Table 1).

Table 2 shows the hyperparameters of the scaled \( \chi^2 \) priors (\( \nu \) and \( S^2 \)) tested for \( \sigma^2_a \) and \( \sigma^2_m \). The values chosen for these hyperparameters generated a wide range of shapes for this distribution that reflected different a priori knowledge on the expected values of both variance components. Model fit and complexity were evaluated with DIC (SPIEGELHALTER et al. 2002), and PR3 was favored (DIC = 3,975.50) although the difference was slight with respect to PR1 and PR4 (DIC = 3,976.79 and 3,976.71, respectively; Table 2). PR2 was penalized with a 2-units greater DIC than PR3, and the model without genetic effects (PR0) was discarded (DIC = 3,981.91). Note that differences in DIC > 3 are generally considered as statistically relevant (BURNHAM and ANDERSON 1998; SPIEGELHALTER et al. 2002), whereas lower discrepancies do
not provide a strong evidence of a better fit and lower degree of model complexity for a given comparison. It is important to highlight that three different MCMC chains were launched for each model and DIC showed a very small variance within models (Table 2).

As shown in Table 3, PR1, PR3 and PR4 models provided very similar estimates of variance components and their ratios, whereas PR2 had a lower value of $\sigma_a^2$. Taking the PR3 model as reference and after correcting for systematic effects (generation and parturition number of the female), the most important source of variation was that due to the uncontrolled factors accounted for $\sigma_e^2$, its mode being 3.842 pups$^2$ (Table 3). Nevertheless, genetic variances $\sigma_a^2$ and $\sigma_m^2$ were high for an inbred strain, 0.151 and 0.035, respectively. It is important to note that the highest posterior density region at 95 % (HPD95) for both variance components was away from zero, starting at 0.066 and 0.017, respectively (Table 3). Mutational heritability was 0.008, and the overall heritability at generation G1 ($h_{G1}^2 = \left(\sigma_a^2 + \sigma_m^2\right) / \left(\sigma_a^2 + \sigma_m^2 + \sigma_p^2 + \sigma_e^2 + \sigma_e^2\right)$) was 0.045 (HPD95 between 0.010 and 0.062), showing that enough additive genetic variance was available to develop a genetic trend under selection or drift. In a similar way, sire and (non-genetic) dam effects had moderate variance components (0.092 and 0.037, respectively), although with wide HPD95 (Table 3).

The C57BL/6J strain did not show noticeable genetic drift during 46 generations. The within-generation average breeding value ($a + m$) ranged between 0 (G1, G2 and G22) and 0.02 pups (G44), with all the HPD95 including the null value (Figure 2). Similarly, within-generation average values for environmental effect ($p_1$ and $p_2$) did not differ from zero. The between-generations phenotypic variability was mainly accounted for the generation number effects (results not shown). Model fit was worse when this systematic effect was dropped out (DIC = 4,018.89).
The within-generation $\sigma_{\tau(i)}^2$ quickly decreased (Figure 3) whereas the new genetic variance originated by mutation increased up to about 0.20 and oscillated around 0.15 thereafter (Figure 4). More specifically, $\sigma_{\tau(i)}^2$ started with a modal estimate of 0.159 in generation G1 (HPD95 between 0.056 and 0.245) and fell to values lower than 0.01 in less than 10 generations. Mutational variance was accumulated during the first generations and reached its maximum at generation G14 (0.231), although it showed an oscillating pattern around 0.15 from generation G7 (HPD95 reached values up to 0.400).

**DISCUSSION**

**Prior distributions and Bayesian analysis:** The mixed model analysis of litter size in C57BL/6J mice was carried out using Bayesian methods. An important characteristic of the Bayesian analysis is that the final inference is based on the posterior distribution, resulting from combining two different sources of information. One of these sources is the experimental data itself and it is not influenced by arbitrary choices other than the model used for analysis. The other source of information is the assignment of prior distributions, which are arbitrarily chosen from previous knowledge of the parameters of interest. If previous information is not available, prior distributions become a blind choice and they could have a substantial impact on the posterior inference (GIANOLA and FERNANDO 1986; BLASCO 2001). Genetic components of litter size in highly inbred mice have not been previously analyzed and we lacked accurate information on the expected values of $\sigma_a^2$ and $\sigma_m^2$. In order to assess influences of priors for both variance components, the analyses performed here made use of very different prior distributions for $\sigma_a^2$ and $\sigma_m^2$, covering the range of mutational heritabilities reviewed by LYNCH (1988) and HOULE et al. (1996) in other traits and species. Model PR0, the one without genetic components, showed
the poorest model fit and DIC substantially decreased when $\sigma_a^2$ and $\sigma_m^2$ were included in the
model. This provided statistical evidence of the presence of additive genetic variance in this
inbred strain. Models PR1 to PR4 showed a similar fit, although the stringent prior for $\sigma_a^2$ and
$\sigma_m^2$ in model PR2 was moderately penalized. It is important to note that posterior inferences
from model PR1, PR3 and PR4 did not differ substantially (Table 3). This reassuring conclusion
indicated that the experimental data had enough information content to override moderate
influences of prior information, and the model performed better under a vague assumption for
$\sigma_a^2$ and $\sigma_m^2$ over the parameter space.

Note that this analysis could also be performed under a frequentist approach by
maximizing the likelihood function through iterative algorithms. These frequentist methods
produce inferences based on the data and the previous knowledge of the distribution of
estimators in the sampling space, without using prior information. As highlighted by BLASCO
(2001), the distribution of the estimator is used for inferences instead of the distribution of the
parameter, which leads to a rather unnatural form of expressing uncertainly about the results of
an experiment. Within the Bayesian context, conceptual simplicity is gained because inferences
are made from probabilities associated with values of the parameter of interest.

**Genetic variability:** Reported estimates of mutational heritability found in the literature
commonly range between $10^{-4}$ and $5 \times 10^{-2}$ (LYNCH 1988; HOULE *et al.* 1996), although this
parameter has never been estimated for litter size in mice. Our estimate fell within this interval
(0.008; Table 3) and was very close to the values reported by KEIGHTLEY and HILL (1992) and
KEIGHTLEY (1998) for body size in mice. Although WRAY’s (1990) approach assumes that
mutations are small and additive, the inclusion of sire and dam environmental effects accounted
for deviations from the infinitesimal model, allowing for a more accurate estimation of $\sigma_m^2$. Within this context, part of the variability accounted for $\sigma_{p_1}^2$ and $\sigma_{p_2}^2$ could have originated from large mutations. Thus, $\sigma_m^2$ likely underestimates all the genetic variation originated by mutation. Moreover, environmental variance could also account for non-additive genetic mutations (ZHANG et al. 2004). Although $\sigma_{p_1}^2$ and $\sigma_{p_2}^2$ modal estimates suggested a greater impact of $\sigma_{p_1}^2$ on litter size, HPD95 were completely overlapping. It is important to note that $a$ and $m$ accounted for additive genetic effects whereas non-additive genetic sources of variation such as inbreeding or heterosis could have substantial impact on mice litter size (BHUVANAKUMAR et al. 1985; HINRICHs et al. 2007). Given the difficulty in accommodating inbreeding and heterosis on the new mutational variability originated at each generation, as well as other non-additive genetic effects (i.e. epistasis or dominance), we restricted the model to pure additive genetic effects and assumed that the remaining non-additive genetic influences were accounted for $p_1$ and $p_2$ or were absorbed by the residual term.

Mutation variance in the mixed model was modeled as the dispersion term associated with random mutation effects arising in each offspring (WRAY 1990). This parameterization has been typically used in mutation experiments (LYNCH 1988; HOULE et al. 1996) and describes the potential effect of mutations in a very short-time interval. Nevertheless, new mutations are accumulated in successive generations, sometimes fixed or removed due to selection or genetic drift, and $\sigma_m^2$ must be viewed in highly inbred strains as a lower limit of the accumulated (mutation) genetic variance. As shown in Figure 4, the within-generation mutation variance increased with generation, up to G14. This is in agreement with a continuous input of new mutations with effects on litter size. The oscillating pattern around 0.15 after G14 agrees with
\( \sigma^2_{u(i)} \) at G1 and suggests an equilibrium between new mutations and the fixation or loss of mutations due to genetic drift and inbred matings. In a similar way, \( \sigma^2_{m} \) lacks additional sources of new variation and its within-generation estimates have a rapid decrease after a few generations, which is related to the small population size and the mating system. Both within-generation \( \sigma^2_{u} \) and \( \sigma^2_{m} \) estimates support the conclusion that this C57BL/6J strain maintained a substantial degree of genetic variance across generations, accounting for an overall heritability around 0.05 (Table 3). This value is clearly lower than the ones reported in outbred mice populations (0.15-0.33; Falconer 1960; Joakimsen and Baker 1977; Long et al. 1991) although it indicates a substantial and generally unaccounted for degree of genetic variability in an inbred strain.

**Absence of environmental and genetic trend across generations:** Mice bred under inbred mating systems for prolonged periods should fix the vast majority (potentially all) of the genetic contribution to variation (Bailey 1982) and typically, individual mice within an inbred strain are considered genetically identical. Nevertheless, unexpected genetic variability has been observed in highly inbred mouse strains (Keightley and Hill 1992), even allowing for genetic drift (Bailey 1977) and genetic trend on phenotypic traits (Festing 1973; Keightley 1998). Indeed, incongruities between genetic homogeneity and phenotypic variability were first recognized more than 40 years ago (Wolff 1961) and this issue is an area of concern in laboratory species. The C57BL/6J strain was developed in the early twentieth century from a very small founder population and it is considered a classical inbred strain with an almost homozygous genome (Wade and Daly 2005). As reported in other inbred mice strains (Falconer 1960), our strain showed substantial phenotypic variation for average litter size.
across generations, average litter size being similar to the estimates reported by other authors (Kirkpatrick et al. 1998; Corva et al. 2004). Note that this population was not under selection for litter size.

Besides the continuous generation of additional genetic variability in our C57BL/6J population, genetic drift did not take place (Figure 2) and changes in the (within-generation) average environmental effect were also negligible. As mentioned above, our mixed model analysis included generation number as a systematic effect and it accounted for the major differences between generations on litter size. Note that model fit was worse when this effect was dropped out (DIC = 4,018.89). Theoretically, this effect must be viewed as the generation-specific contribution of multiple environmental sources of variation (e.g. food, housing or management among others) although some genetic contributions could be involved too. Additive genetic variability is accounted for a and m, and residual effects can absorb (individual-specific) non-additive genetic influences. Nevertheless, a small number of breeding individuals contribute to the next generation and commonly, they are closely related (full-sibs in the majority of cases). Non-additive genetic effects from a given ancestor can be widely spread in the following generation, reducing between-individuals variability and therefore, being partially accounted for the within-generation overall mean (generation number effect). Within this context, genetic drift cannot be completely discarded although, if present, it would be due to non-additive mutations.

In conclusion, we present a new approach to Wray’s (1990) method for modeling $\sigma^2_a$ and $\sigma^2_m$ within a Bayesian framework, where all parameters in the mixed model can be inferred using the Gibbs sampling algorithm. The analysis of litter size in the C57BL/6J strain indicated a low mutational input of genetic variance per generation ($\sigma^2_m = 0.035$), although the accumulation of new mutations in successive generations led to a substantial amount of additive genetic
variability ($h^2 = 0.045$). While genetic uniformity of highly inbred strains is a key point in several research areas (Stevens et al. 2007), our estimates do not support this assumption and confirm a continuous and unavoidable flow of new genetic variability. These results contribute to understanding the mutation/drift equilibrium in experimental populations.

ACKNOWLEDGEMENTS

This research was supported by the National Research Initiative grant number 2005-35205-15453 from the USDA Cooperative State Research, Education, and Extension Service and by the California Agricultural Experimental Station. The research contract of J. Casellas was partially financed by Spain’s Ministerio de Educación y Ciencia (program Juan de la Cierva). The useful comments of Dr. D. Gianola and an anonymous referee are greatly appreciated.

LITERATURE CITED


BRADFORD, G. E., and T. R. FAMULA, 1984 Evidence for a major gene for rapid postweaning

Information Theoretic Approach. Springer-Verlag, New York.

CABALLERO, A., M. A. TORO and C. LÓPEZ-FANJUL, 1991 The response to artificial selection

CORVA, P. M., N. C. MUCCI, K. EVANS, J. F. MEDRANO, 2004 Diet effects on female
reproduction in high growth (hgl/hg) mice that are deficient in the Socs-2 gene. Reprod.
Nutr. Dev. 44, 303-312.

FALCONER, D. S., 1960 The genetics of litter size in mice. J. Cell. Comp. Physiol. 56 (Suppl. 1):
153-167.

FESTING, M., 1973 A multivariate analysis of subline divergence in the shape of the mandible in

FISHER, R. A., 1918 The correlation between relatives under the supposition of Mendelian

Sci. 63, 217-244.

GIANOla, D., R. L. FERNando, S. IM, and J. L. FOULLEY, 1989 Likelihood estimation of
quantitative genetic parameters when selection occurs: models and problems. Genome
31, 768-777.

Carlo, pp. 1-19 in Markov Chain Monte Carlo in Practice, edited by W. R. GILOKS, S.
RICHARDSON and D. J. SPIEGELHALTER. Capman & Hall, London.


**APPENDIX**

Following in part Wray (1990), the additive genetic (co)variance matrix including mutation effects can be partitioned as follows:

\[
A_0\sigma_a^2 + M\sigma_m^2 = A_0\sigma_a^2 + \sum_{k=0}^{t} A_k\sigma_m^2
\]
where \( t \) is the number of generations, \( \mathbf{A}_k \) is the full relationship matrix of additive genetic effects attributed to mutations arising in time unit (or generation) \( k \), and \( \mathbf{A}_0 \) if the full relationship matrix including all individuals in the pedigree. For the mixed model equations, both \( \mathbf{A}_0^{-1} \) and \( \mathbf{M}^{-1} \) are required, computational efficiency becoming a key point. As developed by Quaas (1976), \( \mathbf{A}_0^{-1} \) can be recursively computed from a list of individual, sire and dam identifications ordered by age of individuals. Following in part Wray (1990), \( \mathbf{M}^{-1} \) can also be computed from an age-ordered pedigree with \( n \) individuals and three vectors, \( \mathbf{u} \), \( \mathbf{v} \) and \( \mathbf{h} \), all with dimension \( n \times 1 \). Computation efficiency is gained with \( n \) rounds, with the following calculations in the \( i \)th round:

\[
v_i = \begin{cases} 
\sqrt{\frac{u_p + u_q - h_p + h_q}{4} + 1} & \text{both parents of } i \text{ are known (} p \text{ and } q) \\
\sqrt{\frac{u_q - h_q}{4} + \frac{1}{2}} & \text{only one parent of } i \text{ is known (} q) \\
1 & \text{neither parent of } i \text{ is known}
\end{cases}
\]

where \( v_i \), \( u_i \) and \( h_i \) are the \( i \)th element in vectors \( \mathbf{v} \), \( \mathbf{u} \) and \( \mathbf{h} \), respectively. For \( j = i + 1, \ldots, n \),

\[
v_j = \begin{cases} 
v_p + v_q \quad & \text{if } i \leq p_j < q_j \\
\frac{v_p}{2} \quad & \text{if } j < i \leq q_j \\
\frac{v_q}{2} \quad & \text{if } p_j < i \leq q_j \\
0 \quad & \text{if } p_j \leq q_j < i
\end{cases}
\]

where \( p_j \) and \( q_j \) are parents of the \( j \)th individual and \( p_j < q_j \). For \( j = i + 1, \ldots, n \),

\[
h_j = \begin{cases} 
h_j + \frac{v_p v_q}{2} \quad & \text{if } i \leq p_j < q_j \\
h_j \quad & \text{if } p_j < i
\end{cases}
\]
and for \( j = i, \ldots, n \),

\[
  u_j = u_j + v_j^2
\]

To construct \( M^{-1} \), simply add:

\[ a) \] If both parents of \( i \) are known (\( p \) and \( q \))

- \( v_i^{-2} \) to \( m_{ii} \)
- \( -v_i^{-2}/2 \) to \( m_{ip} \), \( m_{qi} \), \( m_{iq} \) and \( m_{qq} \)
- \( v_i^{-2}/4 \) to \( m_{pp} \), \( m_{pq} \), \( m_{qp} \) and \( m_{qq} \)

where \( m_{kl} \) is the element in the \( k \)th row and \( l \)th column of \( M^{-1} \).

\[ b) \] If only one parent of \( i \) is known (\( q \))

- \( v_i^{-2} \) to \( m_{ii} \)
- \( -v_i^{-2}/2 \) to \( m_{iq} \) and \( m_{qi} \)
- \( v_i^{-2}/4 \) to \( m_{qq} \)

\[ c) \] If neither parent of \( i \) is known.

- \( v_i^{-2} \) to \( m_{ii} \)

Note that this construction of \( A_0^{-1} \) and \( M^{-1} \) allows for a separate inference of additive genetic effects in the base population and new mutations arising in successive generations, \( A_0^{-1} \) and \( M^{-1} \) being independent from the remaining parameters in model. If Bayesian mixed models are applied, this parameterization leads to well known conditional posterior distributions for all parameters, allowing for standard Gibbs sampling. On the contrary, conditional posterior distributions for several parameters under the original Wray’s (1990) approach do not lead to known densities, requiring generic sampling processes (i.e. Metropolis-Hastings sampling) and
the intensive reconstruction of the inverse of the relationship matrix within each sampling iteration (WRAY 1990).
Table 1.- Number of males and females mated per generation (in parenthesis, contributors to the next generation), number of litters and average litter size per generation.

<table>
<thead>
<tr>
<th>Generation</th>
<th>Males</th>
<th>Females</th>
<th>Litters</th>
<th>Mean ± SE</th>
<th>Generation</th>
<th>Males</th>
<th>Females</th>
<th>Litters</th>
<th>Mean ± SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>2 (2)</td>
<td>6 (5)</td>
<td>19</td>
<td>9.42 ± 0.29</td>
<td>G27</td>
<td>10 (5)</td>
<td>18 (6)</td>
<td>18</td>
<td>7.72 ± 0.38</td>
</tr>
<tr>
<td>G2</td>
<td>12 (4)</td>
<td>12 (4)</td>
<td>12</td>
<td>6.50 ± 0.48</td>
<td>G28</td>
<td>12 (3)</td>
<td>19 (4)</td>
<td>30</td>
<td>7.13 ± 0.40</td>
</tr>
<tr>
<td>G3</td>
<td>5 (2)</td>
<td>5 (2)</td>
<td>5</td>
<td>8.40 ± 0.60</td>
<td>G29</td>
<td>11 (3)</td>
<td>19 (4)</td>
<td>24</td>
<td>7.33 ± 0.40</td>
</tr>
<tr>
<td>G4</td>
<td>5 (1)</td>
<td>5 (1)</td>
<td>9</td>
<td>6.55 ± 0.60</td>
<td>G30</td>
<td>6 (3)</td>
<td>12 (4)</td>
<td>17</td>
<td>7.47 ± 0.52</td>
</tr>
<tr>
<td>G5</td>
<td>4 (2)</td>
<td>5 (2)</td>
<td>7</td>
<td>8.42 ± 0.64</td>
<td>G31</td>
<td>8 (4)</td>
<td>11 (5)</td>
<td>14</td>
<td>5.78 ± 0.40</td>
</tr>
<tr>
<td>G6</td>
<td>4 (4)</td>
<td>4 (4)</td>
<td>11</td>
<td>8.18 ± 0.81</td>
<td>G32</td>
<td>6 (4)</td>
<td>13 (5)</td>
<td>16</td>
<td>5.81 ± 0.48</td>
</tr>
<tr>
<td>G7</td>
<td>7 (5)</td>
<td>7 (5)</td>
<td>8</td>
<td>7.12 ± 0.83</td>
<td>G33</td>
<td>6 (6)</td>
<td>10 (10)</td>
<td>16</td>
<td>6.75 ± 0.42</td>
</tr>
<tr>
<td>G8</td>
<td>3 (3)</td>
<td>5 (4)</td>
<td>6</td>
<td>7.66 ± 0.55</td>
<td>G34</td>
<td>14 (6)</td>
<td>23 (6)</td>
<td>41</td>
<td>6.04 ± 0.44</td>
</tr>
<tr>
<td>G9</td>
<td>5 (4)</td>
<td>7 (5)</td>
<td>9</td>
<td>7.88 ± 0.51</td>
<td>G35</td>
<td>7 (4)</td>
<td>12 (6)</td>
<td>27</td>
<td>5.92 ± 0.51</td>
</tr>
<tr>
<td>G10</td>
<td>7 (6)</td>
<td>14 (8)</td>
<td>17</td>
<td>7.94 ± 0.34</td>
<td>G36</td>
<td>11 (6)</td>
<td>19 (8)</td>
<td>31</td>
<td>6.90 ± 0.49</td>
</tr>
<tr>
<td>G11</td>
<td>7 (6)</td>
<td>9 (7)</td>
<td>9</td>
<td>7.55 ± 0.29</td>
<td>G37</td>
<td>6 (5)</td>
<td>14 (8)</td>
<td>32</td>
<td>6.53 ± 0.40</td>
</tr>
<tr>
<td>G12</td>
<td>9 (6)</td>
<td>13 (7)</td>
<td>16</td>
<td>8.12 ± 0.44</td>
<td>G38</td>
<td>10 (5)</td>
<td>23 (9)</td>
<td>47</td>
<td>6.95 ± 0.34</td>
</tr>
<tr>
<td>G13</td>
<td>10 (7)</td>
<td>14 (9)</td>
<td>17</td>
<td>6.64 ± 0.63</td>
<td>G39</td>
<td>12 (7)</td>
<td>27 (9)</td>
<td>35</td>
<td>6.82 ± 0.36</td>
</tr>
<tr>
<td>G14</td>
<td>9 (6)</td>
<td>15 (9)</td>
<td>25</td>
<td>7.52 ± 0.44</td>
<td>G40</td>
<td>13 (8)</td>
<td>32 (9)</td>
<td>39</td>
<td>6.33 ± 0.36</td>
</tr>
<tr>
<td>G15</td>
<td>9 (5)</td>
<td>19 (7)</td>
<td>19</td>
<td>7.73 ± 0.35</td>
<td>G41</td>
<td>10 (8)</td>
<td>20 (11)</td>
<td>20</td>
<td>7.10 ± 0.23</td>
</tr>
<tr>
<td>G16</td>
<td>8 (3)</td>
<td>15 (6)</td>
<td>15</td>
<td>7.40 ± 0.48</td>
<td>G42</td>
<td>12 (4)</td>
<td>28 (6)</td>
<td>33</td>
<td>8.30 ± 0.33</td>
</tr>
<tr>
<td>G17</td>
<td>9 (4)</td>
<td>18 (5)</td>
<td>24</td>
<td>7.29 ± 0.39</td>
<td>G43</td>
<td>7 (2)</td>
<td>18 (3)</td>
<td>43</td>
<td>8.55 ± 0.30</td>
</tr>
<tr>
<td>G18</td>
<td>6 (3)</td>
<td>10 (3)</td>
<td>17</td>
<td>7.70 ± 0.45</td>
<td>G44</td>
<td>2 (1)</td>
<td>8 (2)</td>
<td>26</td>
<td>8.38 ± 0.45</td>
</tr>
<tr>
<td>G19</td>
<td>4 (4)</td>
<td>8 (7)</td>
<td>18</td>
<td>7.16 ± 0.50</td>
<td>G45</td>
<td>4 (4)</td>
<td>11 (6)</td>
<td>18</td>
<td>8.61 ± 0.30</td>
</tr>
<tr>
<td>G20</td>
<td>13 (11)</td>
<td>19 (15)</td>
<td>36</td>
<td>7.66 ± 0.31</td>
<td>G46</td>
<td>6</td>
<td>16</td>
<td>16</td>
<td>7.93 ± 0.29</td>
</tr>
<tr>
<td>G21</td>
<td>28 (23)</td>
<td>49 (30)</td>
<td>79</td>
<td>7.86 ± 0.20</td>
<td>G22b</td>
<td>49 (16)</td>
<td>96 (31)</td>
<td>239</td>
<td>7.96 ± 0.13</td>
</tr>
<tr>
<td>G22</td>
<td>10 (5)</td>
<td>16 (5)</td>
<td>16</td>
<td>6.68 ± 0.45</td>
<td>G23b</td>
<td>40 (24)</td>
<td>93 (40)</td>
<td>221</td>
<td>7.55 ± 0.14</td>
</tr>
<tr>
<td>G23</td>
<td>7 (5)</td>
<td>9 (6)</td>
<td>14</td>
<td>6.78 ± 0.56</td>
<td>G24b</td>
<td>43 (22)</td>
<td>99 (37)</td>
<td>220</td>
<td>8.07 ± 0.14</td>
</tr>
<tr>
<td>G24</td>
<td>7 (2)</td>
<td>13 (2)</td>
<td>24</td>
<td>7.29 ± 0.48</td>
<td>G25b</td>
<td>36 (17)</td>
<td>93 (27)</td>
<td>187</td>
<td>7.59 ± 0.17</td>
</tr>
<tr>
<td>G25</td>
<td>4 (4)</td>
<td>5 (4)</td>
<td>5</td>
<td>6.40 ± 1.07</td>
<td>G26b</td>
<td>33</td>
<td>72</td>
<td>126</td>
<td>7.94 ± 0.20</td>
</tr>
<tr>
<td>G26</td>
<td>4 (4)</td>
<td>8 (6)</td>
<td>13</td>
<td>7.61 ± 0.52</td>
<td>Overall</td>
<td>572 (298)</td>
<td>1,116 (424)</td>
<td>1,986</td>
<td>7.58 ± 0.05</td>
</tr>
</tbody>
</table>
Table 2.- $\chi^2$ hyperparameters specifications and deviance information criterion (DIC) estimates.

<table>
<thead>
<tr>
<th>$\chi^2$ hyperparameters</th>
<th>PR0</th>
<th>PR1</th>
<th>PR2</th>
<th>PR3</th>
<th>PR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu$</td>
<td>-2</td>
<td>1</td>
<td>10</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>$S^2$</td>
<td>0</td>
<td>0.02</td>
<td>0.05</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chain 1</td>
<td>3,981.92</td>
<td>3,976.65</td>
<td>3,977.76</td>
<td>3,975.51</td>
<td>3,976.62</td>
</tr>
<tr>
<td>Chain 2</td>
<td>3,981.98</td>
<td>3,976.86</td>
<td>3,977.74</td>
<td>3,975.52</td>
<td>3,976.72</td>
</tr>
<tr>
<td>Chain 3</td>
<td>3,981.83</td>
<td>3,976.86</td>
<td>3,977.74</td>
<td>3,975.47</td>
<td>3,976.78</td>
</tr>
<tr>
<td>Mean</td>
<td>3,981.91</td>
<td>3,976.79</td>
<td>3,977.75</td>
<td>3,975.50</td>
<td>3,976.71</td>
</tr>
<tr>
<td>SD</td>
<td>0.08</td>
<td>0.12</td>
<td>0.01</td>
<td>0.03</td>
<td>0.08</td>
</tr>
</tbody>
</table>
Table 3.- Modal estimates (and highest posterior density region at 95%) for the variance components and heritabilities.

<table>
<thead>
<tr>
<th>Model (depending on priors for $\sigma^2_a$ and $\sigma^2_m$)</th>
<th>PR0</th>
<th>PR1</th>
<th>PR2</th>
<th>PR3</th>
<th>PR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_a$</td>
<td>0.155</td>
<td>0.020</td>
<td>0.151</td>
<td>0.158</td>
<td></td>
</tr>
<tr>
<td>(0.067 to 0.260)</td>
<td>0.067 to 0.103</td>
<td>0.066 to 0.254</td>
<td>0.070 to 0.273</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_m$</td>
<td>0.033</td>
<td>0.025</td>
<td>0.035</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td>(0.015 to 0.048)</td>
<td>0.009 to 0.043</td>
<td>0.017 to 0.049</td>
<td>0.019 to 0.050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_p_1$</td>
<td>0.112</td>
<td>0.105</td>
<td>0.099</td>
<td>0.099</td>
<td></td>
</tr>
<tr>
<td>(0.011 to 0.683)</td>
<td>0.006 to 0.650</td>
<td>0.003 to 0.604</td>
<td>0.005 to 0.634</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_p_2$</td>
<td>0.051</td>
<td>0.038</td>
<td>0.037</td>
<td>0.037</td>
<td></td>
</tr>
<tr>
<td>(0.009 to 0.612)</td>
<td>0.006 to 0.555</td>
<td>0.002 to 0.596</td>
<td>0.003 to 0.606</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_e$</td>
<td>3.901</td>
<td>3.887</td>
<td>3.842</td>
<td>3.880</td>
<td></td>
</tr>
<tr>
<td>(3.332 to 4.054)</td>
<td>3.326 to 4.101</td>
<td>3.245 to 3.991</td>
<td>3.322 to 3.995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h^2_m$</td>
<td>0.008</td>
<td>0.006</td>
<td>0.008</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>(0.003 to 0.012)</td>
<td>0.002 to 0.009</td>
<td>0.004 to 0.012</td>
<td>0.004 to 0.011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h^2_G_1$</td>
<td>0.045</td>
<td>0.011</td>
<td>0.045</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td>(0.011 to 0.064)</td>
<td>0.002 to 0.046</td>
<td>0.010 to 0.062</td>
<td>0.013 to 0.068</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$$h^2_m = \frac{\sigma^2_a}{(\sigma^2_a + \sigma^2_m + \sigma^2_p_1 + \sigma^2_p_2 + \sigma^2_e)}; \quad h^2_G_1 = \frac{\sigma^2_a + \sigma^2_m}{(\sigma^2_a + \sigma^2_m + \sigma^2_p_1 + \sigma^2_p_2 + \sigma^2_e)}$$
Figure 1.- A priori distributions for $\sigma^2$ and $\sigma^2_m$.

- $\nu = -2; S^2 = 0$
- $\nu = 1; S^2 = 0.02$
- $\nu = 10; S^2 = 0.05$
- $\nu = 0.05; S^2 = 1$
Figure 2.- Mode (black point) and highest posterior density region at 95 % (whiskers) of the average breeding value per generation.
Figure 3.- Mode (black point) and highest posterior density region (whiskers) of the within-generation additive genetic variance ($\sigma^2_{a(t)}$).
Figure 4.- Mode (black point) and highest posterior density region (whiskers) of the within-generation mutational variance ($\sigma^2_{m(i)}$).