

Sensitivity of segregation analysis to data structure and transformation: a case study of trypanotolerance in mice

P. UIMARI*†, S. J. KEMP‡§, J. C. M. DEKKERS†, A. J. TEALE‡ & B. W. KENNEDY†

†Centre for Genetic Improvement, Department of Animal and Poultry Science, University of Guelph, Guelph, Ontario, Canada N1G 2W1 and ‡International Livestock Research Institute, PO Box 30709, Nairobi, Kenya

Sensitivity of segregation analysis for data structure and data transformation was studied using data from two trials in which mice were challenged at three months of age with a cloned isolate of *Trypanosoma congolense* and survival time was recorded. Data included records from three inbred strains (C57BL/6 (tolerant), A/J, and BALB/c (both susceptible)) and their crosses. Data were standardized and normalized using a modified power transformation. Segregation analysis was applied to both untransformed and transformed data to determine the genetic inheritance of trypanotolerance in these mice. Data from the two trials were analysed separately and combined. Four genetic models were compared; a one locus model, a polygenic model, a mixed model with common variance, and a mixed model with different variances for each major genotype. Even though the separate data sets and the combined data set all supported the hypothesis of a major gene (or a tightly linked cluster of genes) with different variances within each genotype, parameter estimates were highly sensitive to data transformation and several sets of parameter estimates gave similar likelihood values because of high dependency between parameters. Based on the results segregation analysis can be very sensitive to data structure in a crossbreeding design and to data transformation. Interpretation of the results can be misleading if the entire parameter space is not studied carefully.

Keywords: mouse, segregation analysis, sensitivity, trypanotolerance.

Introduction

Segregation analysis can be used to determine the genetic inheritance of a particular trait in crossbreeding design (Elston & Stewart, 1973). This is carried out in a maximum likelihood framework where a likelihood ratio test is used to compare different genetic models; a polygenic model (inheritance is explained by many loci with small effects), a monogenic model (inheritance is explained by a locus with a major effect on a trait), and a mixed model (a combination of polygenic and monogenic models) (Elston & Stewart, 1973).

Detection of a major gene using data from a crossbreeding experiment is mainly based on the potential mixture of distributions in F₂ and backcross generations. Thus, results from segregation analysis depend both on how much information is

obtained from the F₂ and backcross generations and on the model fitted in the data; with suboptimal data structure, dependency between parameters of the model fitted in the data can be high.

Another factor that has to be taken into account in segregation analysis is the skewness of the data which can cause false detection of a major gene (MacLean *et al.*, 1975; Go *et al.*, 1978; Demenais *et al.*, 1986). Removing all skewness, however, can lead to considerable reduction in power to detect an existing major gene (Demenais *et al.*, 1986). Thus, the optimum solution for skewed data is to make the transformation simultaneously with estimation of other parameters (MacLean *et al.*, 1984).

The trait studied here is trypanotolerance in mice. The genus *Trypanosoma* includes important pathogenic parasites in humans and domestic animals. One of these is *Trypanosoma congolense* which, among other *Trypanosoma* species, causes Nagana disease in livestock in Africa. It has long been known that certain African cattle breeds, namely N'Dama and West African Shorthorn, are more

*Correspondence. E-mail: puimari@aps.uoguelph.ca

§Present address: Department of Genetics and Microbiology, University of Liverpool, Liverpool L69 3BX, U.K.

tolerant to *Trypanosoma* infection than other breeds (reviewed by Murray *et al.*, 1982). Those breeds are commonly called trypanotolerant and the trait is called trypanotolerance. Trypanotolerance has been found to be associated with the animal's ability to control parasitaemia and to resist anaemia (Paling *et al.*, 1991).

Although the ultimate goal is to isolate genes (or markers linked to those genes) that control trypanotolerance in cattle (Teale, 1993), laboratory studies with mice can help in the search for candidate tolerance genes and give important information on host-parasite interactions. In laboratory experiments, mice are usually challenged by injection and tolerance is measured either as survival time after challenge or as level of parasitaemia. Many laboratory studies have shown that some inbred mouse strains have better trypanotolerance than others. For example, the C57BL/6 strain has better tolerance to *T. congolense* than the A/J and BALB/c strains, which are among the most susceptible strains (Morrison *et al.*, 1978). However, the genetic nature of the tolerance is not known. It has been proposed that survival time is under complex genetic control (Morrison & Murray, 1979) and that the level of parasitaemia is controlled by a single autosomal gene (Pinder, 1984).

The main objective of this paper is to study the sensitivity of the segregation analysis for different combinations of data if data are or are not transformed to adjust for non-normality. The goal of the segregation analysis applied is to determine the mode of inheritance of survival time after *T. congolense* challenge in mice using three inbred strains and their crosses. More specifically, the hypothesis of one gene (or a tightly linked group of genes) explaining a major part of trypanotolerance is tested and the effect of the potential major gene and residual variances within each of its genotypes are estimated.

Materials and methods

Data were obtained from two trials which were carried out at the International Livestock Research Institute (ILRI, former ILRAD), Nairobi, Kenya. The experiment was originally designed for linkage analysis using an approach by Darvasi & Soller (1992); however, to have a prior knowledge of the inheritance of the trait studied a segregation analysis without a marker information was performed. In the first trial, BALB/c (susceptible) and C57BL/6 (less susceptible) inbred mouse strains and their crosses were challenged. For data from this experiment

(data 1) the following notation will be used for the respective strains and crosses: BALB for BALB/c, C57 for C57BL/6, BC_{BC} for backcross of F₁ to BALB and F_{2BC} for the F₂ generation. Data from the second trial (data 2) involved A/J (susceptible) and C57 inbred strains and their crosses (F_{1AC} and F_{2AC}). Mice were challenged at three months of age with a cloned isolate of *T. congolense*. Parasitaemia was checked from a tail blood sample by a thick film smear until they were confirmed positive, and survival time was recorded.

Mice that did not die before day 322 (data 1) or day 264 (data 2) were killed and were not included in the analyses. In total 17 out of 815 mice survived the first trial (six C57, one BC_{BC}, and 10 F_{2BC}), and 24 out of 1020 survived the second trial (six C57, one F_{1AC}, and 17 F_{2AC}). The original data had equal numbers of females and males. Descriptive statistics of the data are given in Table 1 and distributions of the data in Fig. 1.

The two data sets were analysed separately and jointly. C57 mice linked the two data sets together. Combining the data sets was possible because mean survival times of the C57 strain were not statistically different (using a *t*-test) in the two trials and because management and challenge of mice was identical for the two trials.

Segregation analysis (Elston & Stewart, 1973) was used to fit four genetic models to the data: a one locus model, a polygenic model, a mixed model with common variance within major genotypes, and a mixed model with different variances within major genotypes. The log-likelihood function in its most general form, i.e. the mixed model with different variances for the combined data, was:

Table 1 Number of observations (*N*), mean and standard deviation (SD) of survival time (days) of different strains of mice and their crosses

Genetic group	<i>N</i>	Mean	SD
Data 1:			
BALB/c	20	73	16
C57BL/6	14	121	56
BC _{BC} †	375	86	26
F _{2BC}	389	93	33
Data 2:			
A/J	55	52	15
C57BL/6	49	104	30
F _{1AC}	54	141	23
F _{2AC}	838	102	33

†Backcross (BALB/c × C57BL/6) × BALB/c.

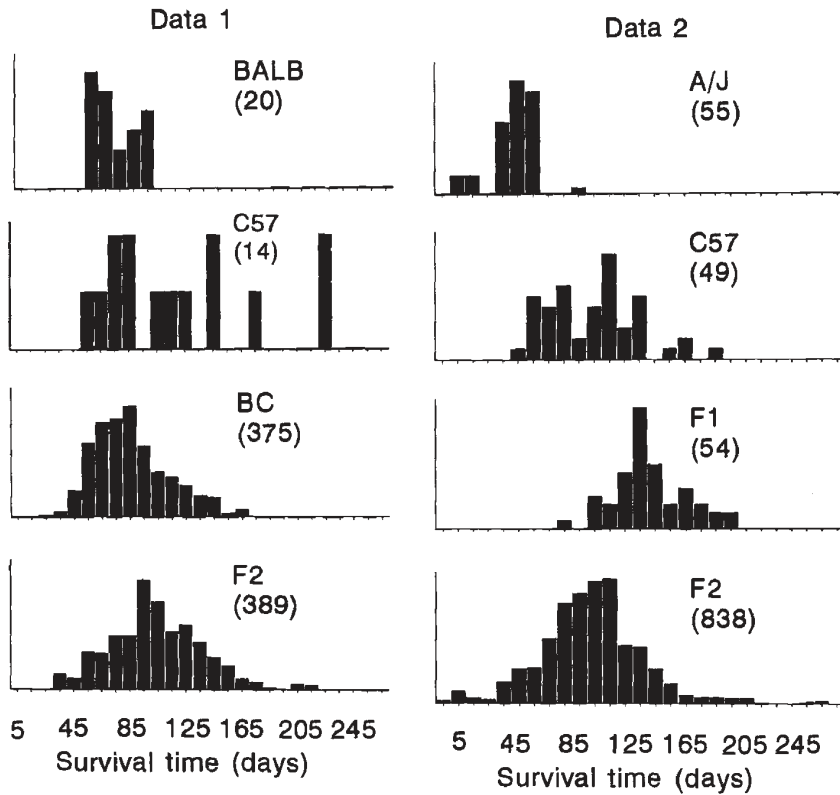


Fig. 1 Relative frequency (percentages) of survival time (number of animals in parentheses). Each bar represents a 10-day period.

$$\begin{aligned}
 \log L(y) = & \sum_{i=1}^{n_1} \log f(y_{1i} - AA - a, \sigma_{AA}^2) \\
 & + \sum_{i=1}^{n_2} \log f(y_{2i} - BB - b, \sigma_{BB}^2) \\
 & + \sum_{i=1}^{n_3} \log f(y_{3i} - CC - c, \sigma_{CC}^2) \\
 & + \sum_{i=1}^{n_4} \log f\left(y_{4i} - AC - \frac{1}{2}(a+c) - h_{AC}, \sigma_{AC}^2\right) \\
 & + \sum_{i=1}^{n_5} \log \left\{ \frac{1}{2} f\left(y_{5i} - BB - \frac{3}{4}b - \frac{1}{4}c - \frac{1}{2}h_{BC}, \sigma_{BB}^2\right) \right. \\
 & \quad \left. + \frac{1}{2} f\left(y_{5i} - BC - \frac{3}{4}b - \frac{1}{4}c - \frac{1}{2}h_{BC}, \sigma_{BC}^2\right) \right\} \\
 & + \sum_{i=1}^{n_6} \log \left\{ \frac{1}{4} f\left(y_{6i} - AA - \frac{1}{2}(a+c+h_{AC}), \sigma_{AA}^2\right) \right. \\
 & \quad \left. + \frac{1}{2} f\left(y_{6i} - AC - \frac{1}{2}(a+c+h_{AC}), \sigma_{AC}^2\right) \right\}
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{4} f\left(y_{6i} - CC - \frac{1}{2}(a+c+h_{AC}), \sigma_{CC}^2\right) \Big\} \\
 & + \sum_{i=1}^{n_7} \log \left\{ \frac{1}{4} f\left(y_{7i} - BB - \frac{1}{2}(b+c+h_{BC}), \sigma_{BB}^2\right) \right. \\
 & \quad + \frac{1}{2} f\left(y_{7i} - BC - \frac{1}{2}(b+c+h_{BC}), \sigma_{BC}^2\right) \\
 & \quad \left. + \frac{1}{4} f\left(y_{7i} - CC - \frac{1}{2}(b+c+h_{BC}), \sigma_{CC}^2\right) \right\}
 \end{aligned}$$

where

$$f(y - \alpha, \sigma_k^2) = \frac{1}{\sqrt{2\pi} \sigma_k} e^{-\frac{1}{2} \left(\frac{y - \alpha}{\sigma_k}\right)^2}$$

and *AA*, *BB*, *CC*, *AC*, *BC* are effects of major genotypes for A/J, BALB, C57, F1_{AC}, and F1_{BC}, respectively. The polygenic component was divided into line effects (*a*, *b* and *c* for inbred strains A/J, BALB and C57, respectively) and heterosis effects (*h_{AC}* and *h_{BC}*). Subscripts 1–7 correspond to A/J, BALB, C57, F1_{AC}, BC_{BC}, F2_{AC} and F2_{BC}, respectively. Residual

variances for different genotypes are σ_k^2 , where k is one of the major genotypes. Restricting any of the parameters leads to a submodel. For example, equal variances for all genotypes (σ_k^2), and restricting $AA = BB = CC = AC = BC$ gives the polygenic model.

The first term in the log-likelihood function describes the joint probability density distribution for the n_1 observations from the inbred A/J strain to be normal with mean equal to the major gene effect AA plus the line effect a , and variance σ_{AA}^2 . Similarly, the next three terms characterize normal distributions for the n_2, n_3 and n_4 observations from the inbred BALB and C57 strains and from the $F1_{AC}$, respectively. The fifth term describes that the joint probability density distribution of n_5 observations for the BC_{BC} is a mixture of two normal distributions with equal mixing proportions, i.e. a mouse from BC_{BC} has equal probability of having either major gene BB or BC . Analogously, for the F_2 generations, a mouse has a probability of $\frac{1}{4}$ having either of the homozygous parental genotypes and a probability of $\frac{1}{2}$ of being heterozygous (sixth and seventh terms). Average polygenic components (line effect + heterosis) are the same for all the mice within each of the parental strains and crosses.

To avoid false detection of a major gene because of skewness of the data, data were standardized to mean 0 and variance 1 and normalized by a modified power transformation based on the following formula (MacLean *et al.*, 1984):

$$y = \frac{r}{p} \left(\left(\frac{x}{r} + 1 \right)^p - 1 \right),$$

where x is a standardized observation, y is a transformed observation, r is a constant to ensure that the logarithm is taken from positive values (r was set to 3) and p is the power of the transformation. For example, p equal to 1 means no transformation and p equal to 0.5 is approximately a square root transformation. For p equal to 0 the formula must be modified to:

$$y = r \log \left(\frac{x}{r} + 1 \right).$$

The best transformation was estimated simultaneously with all other parameters of the likelihood function by estimating the maximum likelihood of the model for different p -values varying from 0 to 1 and plotted against p . In order to achieve this, the transformation of variable formula was used to derive a log-likelihood value of the untransformed data (MacLean *et al.*, 1984):

$$\log L(x) = \log L(y) + (p-1) \sum_{i=1}^N \log \left(\frac{x_i}{r} + 1 \right),$$

where $\log L(y)$ is the log-likelihood value of the transformed data and N is the total number of observations in the data. The value corresponding to the highest peak of the likelihood profile indicates the maximum likelihood estimate of p . An alternative way to transform data is given by Elston (1984) where nonsegregating generations (parental and F_1) are first used to find the best transformation of the data. Then all the available data are transformed using the maximum likelihood estimate of p and finally, a segregation analysis is performed with the transformed data. The small number of observations in these generations in our data did not allow the approach suggested by Elston (1984).

The hypothesis of a major gene explaining part of the trypanotolerance was tested using a likelihood ratio test between the polygenic model and the mixed model with common variance. The hypothesis of different variances for each genotype was tested

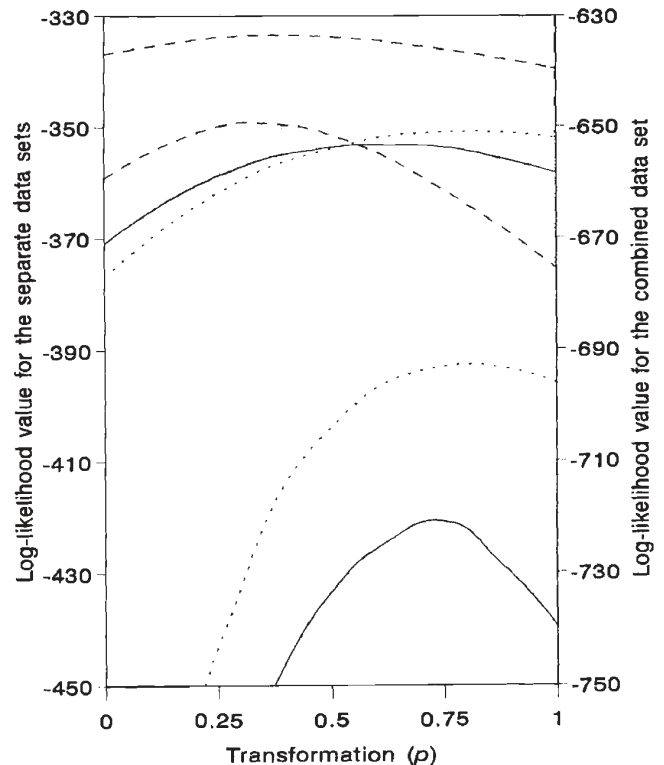


Fig. 2 Log-likelihood values for power transformation (p) using data 1 (broken line), data 2 (dotted line) and the combined data (solid line). For each data set, the higher likelihood profile corresponds to the mixed model with different variances and the lower profile to the mixed model with common variance.

by a likelihood ratio test between the mixed model with common variance and the mixed model with different variances. A χ^2 -distribution with degrees of freedom equal to the difference in the number of estimated parameters between the general model and its submodel was used as the distribution of the likelihood ratio test statistic under the null hypothesis.

Likelihood functions were maximized by Powell's method (Press *et al.*, 1989) and 95 per cent confidence intervals for estimates were obtained by bootstrapping (Efron, 1982). In bootstrapping, 200 data sets were drawn from the original data set by resampling. The new data sets were analysed and parameter estimates were stored and ordered. The 5th and the 195th ordered estimates gave the approximate 95 per cent confidence interval. All estimates were transformed back to the original scale; parameters corresponding to the mean of the distribution had an exact backtransformation and a

first-order Taylor expansion was implemented for variances.

Results

Power transformation

The best transformation (p) for data 1 using the mixed model was 0.375 (Fig. 2) which is stronger than the square root transformation but weaker than the log transformation. For data 2 and for the combined data, the best transformation was 0.75 (Fig. 2), which is weaker than the square root transformation. The best transformation was generally the same for common and different variances. The only exception was the combined data, where 0.75 was best for the mixed model with common variance but 0.625 was best for different variances. However, because the surface was quite flat for the mixed model with different variances and to avoid different

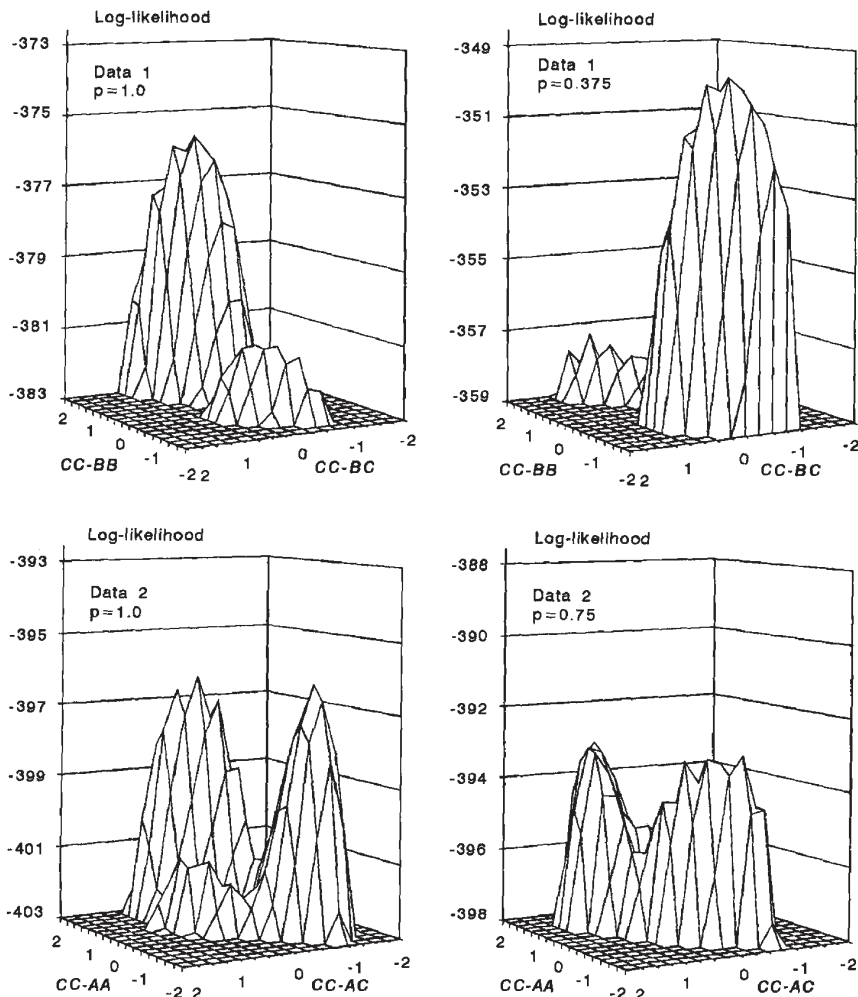


Fig. 3 Log-likelihood surface for data 1 (upper graphs) and for data 2 (lower graphs) without power transformation ($p = 1.0$) and with power transformation ($p = 0.375$ and $p = 0.75$ for data 1 and 2, respectively). Differences between genotypes ($CC-BB$, $CC-BC$, $CC-AA$ and $CC-AC$) are given on the transformed scale; for $p = 1.0$ one unit equals one phenotypic standard deviation on the untransformed scale.

transformations for different mixed models, $p = 0.75$ was used for both models in the case of the combined data.

Likelihood surface

Generally, for each data set and genetic model, the likelihood surface was not unimodal but included several local maxima. This is shown in Fig. 3 for the mixed model with common variance (only values which exceeded the likelihood value of the polygenic model are plotted). For data 1, the maximum of the likelihood function depended on whether the transformation was used. Although the highest likelihood value corresponded to a recessive major gene originating from the C57 strain with or without transformation, the gene had a positive effect on survival time without transformation but a negative effect with transformation. Because data 2 included only a single segregating generation ($F2_{AC}$) the likelihood surface was divided into two sides, with one side being a mirror image of the other. This led to two sets of parameter estimates with the same likelihood value; if a high value was assigned to a major gene, estimates of polygenic effects were low and vice versa.

Hypothesis testing

Maximum likelihood values for different genetic models are given in Table 2. All three data sets (separate and combined data sets) supported rejection

($P < 0.001$) of the null hypothesis of pure polygenic inheritance. Also, the null hypothesis of common variance within genotypes was rejected at the 0.1 per cent level.

Parameter estimation

For the analyses where separate data sets were used only the main results are presented in the text. For the combined data set, estimates of parameters of the mixed model with common variance and with different variances are given in Table 3. Sets of parameters at local maxima with likelihoods close to the global maximum likelihood value are also given in Table 3.

As discussed earlier, the likelihood surface for the mixed model using data 1 had two areas within the parameter space where the maximum likelihood value exceeded the likelihood value of the polygenic model (Fig. 3). Without transformation, the data were explained by a recessive *C*-gene (a gene from the C57 strain) with an effect of 41 days and the polygenic line effect was small ($c - b = -2$ days). Transformation changed the relative magnitude of the two local maximum likelihood areas (Fig. 3) and the survival data were explained by a large difference between line effects ($c - b = 73$ days). In that case the *C*-gene was still recessive, but had a large negative effect on survival time (-41 days).

Similarly, when the mixed model with different variances was used, the gene from the C57 strain

Table 2 Log-likelihood values ($\log L(x)$) for different genetic models for data 1, data 2 and combined data

Genetic model	Data 1†	Data 2‡	Combined data‡
One locus model	-357.7 (4)	-443.0 (4)	-755.6 (6)
Polygenic model	-360.1 (4)	-397.7 (4)	-730.9 (6)
Mixed model with common variance	-349.5 (6)	-392.9 (6)	-720.8 (10)
Mixed model with different variances	-333.3 (8)	-351.5 (8)	-653.7 (14)
Test statistics:			
A major gene§	21.2***	9.6**	20.2***
Different variances within genotypes¶	32.4***	82.8***	134.2***

Degrees of freedom for each model are given within parentheses.

†Power of data transformation 0.375.

‡Power of data transformation 0.75.

§Mixed model with common variance vs. polygenic model.

¶Mixed model with different variances vs. common variance.

Significance levels: *** $P < 0.001$, ** $P < 0.01$.

Table 3 Maximum likelihood estimates† (days) of the mixed model for untransformed ($p = 1.0$) and transformed ($p = 0.75$) combined data

Transformation: Parameter	Common variance					Different variances			
	1.0	0.75	0.75	0.75	0.75	1.0	0.75	0.75	95 per cent confidence interval‡
Line effects§:									
$c-a$	19	12	23	12	23	39	57	31	(14-63)
$c-b$	-7	-7	-7	63	63	3	12	12	(0-34)
Heterosis:									
h_{AC}	75	53	40	53	40	72	77	49	(36-70)
h_{BC}	30	28	28	-8	-8	-4	-10	-9	(-27-6)
Major genotypes§:									
$CC-AA$	37	42	32	42	32	10	-10	16	(-13-30)
$CC-AC$	37	15	-2	15	-2	19	6	-9	(-21-6)
$CC-BB$	39	36	36	-33	-33	24	13	13	(-6-33)
$CC-BC$	39	36	36	-33	-33	-6	-18	-18	(-35-0)
Variances:									
common	803	817	829	823	834				
σ_{AA}^2						236	330	396	(184-519)
σ_{AC}^2						756	544	403	(310-793)
σ_{BB}^2						242	286	286	(227-346)
σ_{BC}^2						720	668	667	(527-877)
σ_{CC}^2						2154	2315	2321	(1858-2679)
Log $L(x)$	-739.6	-720.8	-720.8	-721.0	-721.0	-658.4	-653.7	-654.6	

†Results also shown for several local maxima with likelihood values close to global maximum.

‡Confidence interval for the last column of estimates.

§Estimates for line effects and major genotypes are expressed as a deviation from the line effect c and from the genotype CC .

had a positive effect on survival time without transformation but when the data were transformed the effect was negative. For this model, estimates of variances for different genotypes varied substantially; the variance assigned to the CC genotype was almost 10 times the variance assigned to the BB genotype.

For data 2, when the mixed model with common variance was applied, either the C - or A -gene was estimated to be recessive with an effect of 36 days between homozygotes. Depending on which one had a positive effect, the line effect was either moderate or large. The transformation shifted the likelihood mass to the other side of the surface (Fig. 3), resulting in four sets of estimates with similar likelihood values. Two sets estimated either the C - or A -gene to be a dominant major gene. The other two sets supported a partially dominant gene either from the $C57$ or the A/J strain. The line effect was either large or moderate, depending on which genotype was assigned to have a positive effect on survival time.

When a mixed model with different variances was applied to data 2 without transformation, the maximum likelihood value was achieved when the heterozygous genotype had a negative effect on survival time compared to homozygotes. With transformation, two sets of parameter estimates had almost equal likelihood values. The first set explained survival time by a large polygenic effect and superiority of the F_1 generation caused by heterosis. The second set estimated a moderate polygenic line effect, a major gene from the $C57$ strain, and superiority of the heterozygous genotype. Regardless of transformation, the estimate of variance of the CC genotype was substantially larger than estimates of variances of other genotypes.

With combined data and no transformation, a mode of inheritance where the A -gene was superior over the C -gene was no longer equally as good as a mode of inheritance where the C -gene was superior over the A -gene (Table 3), which was the case when data 2 were analysed separately. Thus, combining the data sets eliminated some of the parameter esti-

mate sets obtained when data 2 were analysed separately, but added an extra one corresponding to data 1 parameters, thus resulting in four sets of parameter estimates with similar likelihood values (Table 3). Two sets estimated the *C*-gene to be dominant over the *A*-gene and recessive over the *B*-gene. The other two sets estimated the *C*-gene to be almost codominant over the *A*-gene and either better or worse than the *B*-gene (Table 3).

When the combined data were analysed using a mixed model with different variances and no transformation was applied (Table 3), estimates were similar to those obtained for the separate data sets. With transformation, two sets of estimates were obtained. Estimates of parameters corresponding to data 2 were similar, but estimates of parameters corresponding to data 1 were different from the estimates from the separate analyses. For the last set of estimates, approximate 95 per cent confidence intervals were computed and are reported in Table 3. Intervals were not symmetric around the maximum likelihood estimates. Also, parameters associated with data 2 had wider confidence intervals than other parameters.

Discussion and conclusions

Even though the hypothesis of a major gene was accepted in all analyses, no single explanation for the mode of inheritance of survival time after *T. congolense* challenge in mice was obtained. Especially, transformation of the data was very critical to estimates. Although, some estimates could be discarded when the two data sets were combined, several possibilities still remain, because the link between the data sets was weak because of the small number of C57 mice. This leaves much for interpretation of the results. For example, even though the mixed model with different variances within genotypes was statistically better than a model with common variance, it is unclear whether this model is biologically more meaningful than assuming common variance. When estimated parameters were fitted to the F_{2AC} generation, the *CC* genotype explained the best and the worst survivors, because of its large variance (Fig. 4). This was not the case for the mixed model with common variance (Fig. 4).

An increase in variances in backcross and F_2 generations resulting from segregation of polygenes was not taken into account, because it would have complicated the analysis considerably (finding the global maximum). However, a model with different variances for parental strains and crosses was applied (results not shown). Although parameter

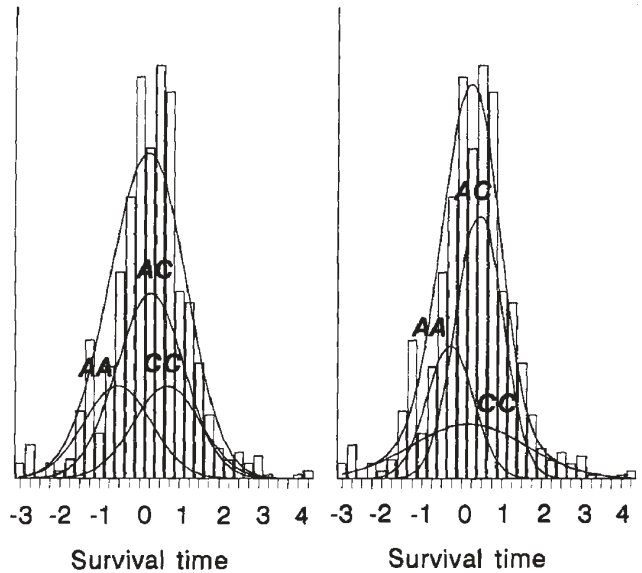


Fig. 4 Fitting the maximum likelihood parameters, estimated from the combined data, to the F_{2AC} data. The left chart is for the mixed model with common variance and the right chart is for the mixed model with different variances for genotypes. Survival time is expressed on a transformed scale ($p = 0.75$). Each bar represents 0.25 units. Curves labelled *AA*, *AC* and *CC* are estimated distributions of mice with *AA*, *AC* and *CC* genotypes, respectively, and the curve without a label is an overall distribution.

estimates from this model were similar to those from the mixed model with common variance, likelihood values of the mixed model with different variances within parental strains and crosses (for data 1 with $p = 0.375$, $\log L(x) = 341.89$; for data 2 with $p = 0.75$, $\log L(x) = 377.12$) were lower than likelihood values of the mixed model with different variances within genotypes (Table 2).

Another important assumption in the analysis was homozygosity of the parental strains. Mouse strains used in this study were from commercial sources that were maintained for decades by full-sib mating. For this reason it is valid to assume that the strains were homozygous. However, a possibility of mutations always exists.

About 2 per cent of mice survived the challenge and were not included in the analyses. However, if arbitrary values had been assigned to those which survived the challenge (for example the last recording day), these observations might have had a substantial effect on results because they were outliers. First, the major gene effect might have been larger, because most of the surviving mice were either from the C57 strain or from the F_2 generations. Secondly, estimates of the variance of the *CC*

genotype might have also been larger, because a larger proportion of C57 mice survived (6 out of 20 in the first trial and 6 out of 55 in the second trial).

In this study, the usual assumption of normal distribution within genotypes was made. This may not be appropriate for survival data, even when data were transformed. It can only be hypothesized that a model with other distributional assumptions could fit data better than a normal distribution. Also, the data from the survival trial might contain aspects that are impossible or at least difficult to take correctly into account in a segregation analysis. These include the periodical nature of the disease (trypanosomiasis), with several peaks of parasitaemia (Morrison *et al.*, 1978). It seems natural to expect that the further the time from the peak of parasitaemia the better the chance is for a mouse to survive. A solution for the drawbacks of the statistical analyses performed in this paper (distributional assumptions, censored data, and underlying biological mechanism) may be the use of generalized linear models developed for survival data (McCullagh & Nelder, 1989). Further research in this area is needed.

In conclusion, segregation analysis was found to be very sensitive to the structure of the crossbreeding data and data transformation. Based on the results it is important to include both backcross and F_2 generation data to be able to reduce the high dependency between parameter estimates and also collect enough data from the parental generations to be able to adjust for non-normality prior to a subsequent segregation analysis. This study clearly shows the importance of the well-known fact that the maximization process has to be started from different points in the parameter space if numerical methods are used to find the maximum of the likelihood function, otherwise some combinations of parameter estimates with equal likelihoods can be missed. In addition to a decrease in the probability of detecting a spurious major gene, transformation of data can lead to a tremendous change in the likelihood surface and thus in parameter estimates. In this study segregation analysis supported the hypothesis of a major gene (or tightly linked group of genes) affecting survival time after *T. congolense* challenge in mice, with different variances within genotypes. However, a multimodal shape of the likelihood surface was observed and different sets of parameter estimates gave similar likelihood values, so that no definite answer for the inheritance of trypanotolerance in mice can be given based on the results reported here.

References

- DARVASI, A. AND SOLLER, M. 1992. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor. Appl. Genet.*, **85**, 353–359.
- DEMEAIS, F., LATHROP, M. AND LALOUEL, J. M. 1986. Robustness and power of the unified model in the analysis of quantitative measurements. *Am. J. Hum. Genet.*, **38**, 228–234.
- EFRON, B. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia.
- ELSTON, R. C. 1984. The genetic analysis of quantitative trait differences between two homozygous lines. *Genetics*, **108**, 733–744.
- ELSTON, R. C. AND STEWART, J. 1973. The analysis of quantitative traits for simple genetic models from parental, F_1 and backcross data. *Genetics*, **73**, 695–711.
- GO, R. C. P., ELSTON, R. C. AND KAPLAN, E. B. 1978. Efficiency and robustness of pedigree segregation analysis. *Am. J. Hum. Genet.*, **30**, 28–37.
- MACLEAN, C. J., MORTON, N. E. AND LEW, R. 1975. Analysis of family resemblance. IV Operational characteristics of segregation analysis. *Am. J. Hum. Genet.*, **27**, 365–384.
- MACLEAN, C. J., MORTON, N. E. AND YEE, S. 1984. Combined analysis of genetic segregation and linkage under an oligogenic model. *Comput. Biomed. Res.*, **17**, 471–480.
- MCCULLAGH, P. AND NELDER, J. A. 1989. *Generalized Linear Models*, 2nd edn. Chapman and Hall, London.
- MORRISON, W. I. AND MURRAY, M. 1979. *Trypanosoma congolense*: Inheritance of susceptibility to infection in inbred strains of mice. *Exp. Parasit.*, **48**, 364–374.
- MORRISON, W. I., ROELANTS, G. E., MAYOR-WITHEY, S. AND MURRAY, M. 1978. Susceptibility of inbred strains of mice to *Trypanosoma congolense*: correlation with changes in spleen lymphocyte populations. *Clin. Exp. Immunol.*, **32**, 25–40.
- MURRAY, M., MORRISON, W. I. AND WHITELAW, D. D. 1982. Host susceptibility to African trypanosomiasis: Trypanotolerance. In: Baker, J. R. and Muller, R. (eds) *Advances in Parasitology*, pp. 1–68. Academic Press, London.
- PALING, R. W., MOLOO, S. K., SCOTT, J. R., GETTINBY, G., MCDIMBA, F. A. AND MURRAY, M. 1991. Susceptibility of N'Dama and Boran cattle to sequential challenges with tsetse-transmitted clones of *Trypanosoma congolense*. *Parasit. Immunol.*, **13**, 427–445.
- PINDER, M. 1984. *Trypanosoma congolense*: Genetic control of resistance to infection in mice. *Exp. Parasit.*, **57**, 185–194.
- PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. AND VETTERLING, W. T. 1989. *Numerical Recipes*. Cambridge University Press New York.
- TEALE, A. 1993. Improving control of livestock diseases. *Bioscience*, **43**, 475–483.