Bayesian Statistical Analyses for Presence of Single Genes Affecting Meat Quality Traits in a Crossed Pig Population

L. L. G. Janss, J. A. M. Van Arendonk and E. W. Brascamp

Department of Animal Breeding, Wageningen Institute of Animal Sciences, Wageningen Agricultural University,
6700 AH Wageningen, The Netherlands

Manuscript received January 2, 1996

Accepted for publication October 14, 1996

ABSTRACT

Presence of single genes affecting meat quality traits was investigated in F_2 individuals of a cross between Chinese Meishan and Western pig lines using phenotypic measurements on 11 traits. A Bayesian approach was used for inference about a mixed model of inheritance, postulating effects of polygenic background genes, action of a biallelic autosomal single gene and various nongenetic effects. Cooking loss, drip loss, two pH measurements, intramuscular fat, shearforce and back-fat thickness were traits found to be likely influenced by a single gene. In all cases, a recessive allele was found, which likely originates from the Meishan breed and is absent in the Western founder lines. By studying associations between genotypes assigned to individuals based on phenotypic measurements for various traits, it was concluded that cooking loss, two pH measurements and possibly backfat thickness are influenced by one gene, and that a second gene influences intramuscular fat and possibly shearforce and drip loss. Statistical findings were supported by demonstrating marked differences in variances of families of fathers inferred as carriers and those inferred as noncarriers. It is concluded that further molecular genetic research effort to map single genes affecting these traits based on the same experimental data has a high probability of success.

SINCE the advent of modern DNA techniques, identification of single genes is receiving increased attention in fundamental and applied sciences. Study of effects of single genes can aid in unravelling physiological processes, which has relevance for many life sciences, and often has relevance across species. For instance, the finding of an obesity gene in mice (ZHANG *et al.* 1994) may have relevance for several other mammals such as humans or pigs. Use of animal populations for identification of single genes can have several advantages unseen in human populations, such as large amounts of data, designed experiments and controlled breeding. This makes the use of animal populations, in this respect, worthy of further attention.

In commercial livestock populations, phenotypic observations are often abundantly available at low costs making preliminary statistical analysis a worthwhile first step in the identification of single genes. Segregation analysis (ELSTON and STEWART 1971; MORTON and MACLEAN 1974) is the most powerful statistical method for identification of single genes (HILL and KNOTT 1990) but, so far, has not found widespread use in animal genetics. Exact computations involved in application of this method are impossible in common situations arising in animal populations, and analytical approximations (*e.g.*, LE ROY *et al.* 1989; KNOTT *et al.* 1992) limit

application to simple models and simple pedigree structures. Recently, however, Gibbs sampling and related Markov chain Monte Carlo methods (GEMAN and GEMAN 1984; GELFAND and SMITH 1990; SMITH and ROBERTS 1993) have been introduced that can facilitate computations in many statistical applications. Gibbs sampling has been used in human genetics for both likelihood-based as well as Bayesian-based inferences in variance component estimation (GUO and THOMPSON 1991), segregation and linkage analysis (GUO and THOMPSON 1992; THOMAS and CORTESSIS 1992), computation of genotype probabilities (*e.g.*, SHEEHAN and THOMAS 1993) and gene mapping (as an example in SMITH and ROBERTS 1993). In animal genetics, Gibbs sampling has been introduced in Bayesian approaches for variance component estimation in linear models (*e.g.*, WANG *et al.* 1993), and in nonlinear models (SORENSEN *et al.* 1995), for analysis of selection response (SORENSEN *et al.* 1994) and for segregation analysis (JANSS *et al.* 1995).

The aim of this paper is to investigate whether single genes may exist that affect meat quality traits, measured in the F_2 of a cross of Chinese Meishan and Western pig lines. Detection of single genes is based on a statistical modelling approach, using a Bayesian approach to segregation analysis described by JANSS *et al.* (1995). The secondary aim of this paper is to demonstrate the flexibility, by virtue of applying Gibbs sampling, of this Bayesian approach, which so far has not been used for the analysis of field data. This highlights details in appli-

Corresponding author: L. L. G. Janss, Department of Animal Breeding, P.O. Box 338, WAU, NL-6700 AH, Wageningen, The Netherlands. E-mail: luc.janss@alg.vf.wau.nl

FIGURE 1.—Design of the crossing experiment to produce F_2 crossbreeds between Chinese Meishan and five Western pig lines. Step 1: 126 F_1 litters were produced from 19 Meishan males and 126 females of five Western lines in five companies. Step 2: 264 F_2 litters were produced from 39 centrally housed F_1 males taken equally from all companies, and from 265 F_1 females having remained in the companies. Step 3: From produced F_2 crossbreeds, animals not used for breeding were centrally slaughtered to measure meat quality traits. All selection steps were random, except selection of nonexperimental F_2 breeding stock.

cation of this new methodology and shows the types of inferences that are produced, which are different from inferences made by classical likelihood-based segregation analysis.

MATERIALS AND METHODS

Data: F_2 crossbreeds between the Chinese Meishan pig breed and Western pig lines were available from an experiment involving five Dutch pig breeding companies (Figure 1). Crossbreeds were produced in two batches at the same time in all companies. For each batch, purebred Western females at the companies were artificially inseminated by a group of 19 purebred Meishan males from a centrally housed population of Meishan animals, producing crossbred F_1 litters. Purebred females at the breeding companies were of Dutch Landrace and Large White types. In total, 126 F_1 crossbred litters were produced. From F_1 litters, a random selection of young males and females were taken as parents to produce F_2 crossbred litters, each female producing one F_2 litter. In total, 264 F_2 litters were produced, descending from 39 F_1 fathers. The 39 F_1 fathers were used across breeding companies through

TABLE 1
Overview of traits

Trait	Full name	<i>N</i>
Drip	Drip Loss (%)	844 2.70 ± 1.54
Cook	Cooking loss (%)	845 26.4 ± 3.46
Shear	Shear force (N)	845 39.6 ± 10.5
Imfat	Intramuscular fat (%)	831 1.84 ± 0.87
pH	pH	845 5.66 ± 0.26
pH-s	pH in <i>M. Semimembranosus</i>	846 5.82 ± 0.30
Light	CIELAB L* color coordinate	844 53.9 ± 4.83
Red	CIELAB a* color coordinate	846 17.3 ± 1.90
Yellow	CIELAB b* color coordinate	845 9.59 ± 1.92
Fat	HGP Back-fat thickness (mm)	846 22.0 ± 5.69
Lean	HGP Back-lean thickness (mm)	844 40.6 ± 6.69

Values are raw means ± SD; *N* is number of observations. Traits were measured in the *M. Longissimus* (loin muscle) except for pH-s, Fat and Lean.

the formation of a central pool and use of artificial insemination; F_1 females remained at the breeding companies where born. This yielded a 75% similar genetic background for all F_2 crossbreeds. From performance-tested F_2 animals, ~1200 in total, ~350 animals were retained for further breeding. The majority of these animals were gilts (>300), chosen at random within the framework of the experiment. The additional animals (<50) were retained by the breeding companies, likely with selection on a combination of production and reproduction traits. Due to the low number involved and emphasis on different traits, the effect of this selection on the traits used in the current analyses is expected to be negligible. Performance tested F_2 animals not retained for breeding were slaughtered at ~90 kg in a central slaughter-house. On these slaughtered animals, several meat quality traits were measured. For genetic analyses, a pedigree file was constructed including F_1 parents and pure line (Meishan and Western) grandparents of the observed F_2 individuals. Presence of Halothane susceptibility (EIKELBOOM and MINKEMA 1974), a common known genetic defect affecting meat quality traits in western lines and that is known as malignant hyperthermia in man, was excluded by molecular typing of pure Meishan founders and F_1 fathers, which were all found free of this Halothane susceptibility mutation. Molecular typing was done by VAN HAERINGEN LABORATORIUM BV (Wageningen, The Netherlands), using methods as described by OTSU *et al.* (1992).

Measurements: In Table 1 numbers of observations, raw means and standard deviations for the traits measured are given. In samples of *M. Longissimus* (loin muscle), pH, drip loss, cooking loss, shearforce, intramuscular fat and color were measured; additionally, a pH measurement was taken in a sample of *M. Semimembranosus* (a ham muscle) and back fat and lean thickness were measured. Color was measured as three coordinates according to the CIELAB L*a*b* system, where L* is a general indication of lightness, a* represents the degree of green-redness and b* represent the degree of blue-yellowness (MACDOUGALL 1986). Fat and lean thickness are based on a single measurement with the Hennessy Grading Probe between the third and fourth rib, 6 cm from the spine, as routinely done in The Netherlands to predict carcass meat percentage. Predicted meat percentages, however, were not analyzed in this study, because the employed prediction equation to predict meat percentage from fat and lean thickness might not hold for the relatively fat Meishan crossbreeds. All traits were measured 24 hr after slaughter, except fat and lean thickness, which were measured directly after slaughter.

Exact details on measurement procedures for these traits can be found in HOVENIER *et al.* (1992). In the following, abbreviations for trait names will be used as given in Table 1.

Nongenetic influences: Meat quality traits can be largely influenced by genetic background and by environment. Well known environmental effects are transport conditions, leading often to large effects of slaughter day or week (CAMERON 1990; HOVENIER *et al.* 1992). Data analyzed here were collected on 26 different slaughter days. In the described data, also an effect of breeding company where the crossbred was produced (five levels) could be expected. This effect could have a partial genetic background, as the maternal granddam of the crossbreds was company specific, and may additionally have a nongenetic basis, *e.g.*, in housing or feeding conditions at the different companies. Because semen of F₁ fathers was exchanged between companies, the genetic basis of a company effect could be separated from nongenetic sources. The design of the experiment was such that also possible effects of breeding company could be separated from effects of slaughter days by slaughtering animals from at least two companies on most of the days. Besides slaughter day and breeding company, sex of the animal (measurements were made on females and on intact males) and its carcass weight were recorded as they may have nongenetic influences on the recorded traits. Significance of these nongenetic effects was investigated by use of a fixed linear model (SAS-GLM, SAS INSTITUTE INC. 1988) fitting slaughter day, breeding company, sex and carcass weight simultaneously. Each variable was found to have a significant effect ($P < 0.01$) on at least several traits; slaughter day was significant for all traits. In further genetic analyses, all nongenetic effects considered were used, thus maintaining for simplicity the same nongenetic effects for analysis of each trait.

Statistical model: A model was used with nongenetic effects of slaughter day, breeding company, sex and carcass weight, and genetic effects of polygenic background genes and a single gene. Polygenic effects were modelled to be strictly additive. The model for the single locus assumed an autosomal biallelic locus with Mendelian transmission probabilities. A possible dominance effect at the single locus was allowed for. Estimation of polygenic variance in the described data will be based on variation between F₂ families, and this will not include possible segregation variance at polygenic loci. Segregation variance refers to the increase of genetic variance that can arise in the F₂ due to allele frequency differences in founder lines (*e.g.*, LANDE 1981). However, assuming the polygenic loci to be large in number and assuming no gene with large effect to be present among the polygenes, segregation variance at polygenes will be negligible. At the single locus, segregation variance was accounted for by modelling of different allele frequencies for the founder groups. The statistical model to describe phenotypic observations on F₂ crossbreds for each trait y is

$$y = X\beta + Zu + ZWm + e \quad (1)$$

In (1), β is a vector of fixed nongenetic effects and X is a design-covariate matrix containing 0/1 dummy variables relating effects of slaughter day, breeding company and sex to observations and containing a column with carcass weights of those individuals with observations in y . Vectors u and Wm contain genetic effects of all individuals in the pedigree considered, which here included genetic effects of F₂ crossbreds, their parents and grandparents. Genetic effects are separated in polygenic effects in u and single-gene effects in Wm . Matrix Z is an incidence matrix relating the genetic effects to observation in y ; Z contains empty columns for individuals without an observation. Vector e contains random errors. Single-gene effects are expressed using W , a four-col-

umn matrix with 0/1 variables to indicate genotypes of individuals, and the vector $m = (-a, d, d, a)'$ that contains the genotypic values. Four genotypes are considered here for notational convenience only; in computations, three genotypes are considered, not distinguishing between the two heterozygotes. In W , the four columns correspond to the possible genotypes denoted as $A_L A_L$, $A_L A_H$, $A_H A_L$ and $A_H A_H$. Allele A_L , with "L" of "Low," is defined as the allele that decreases the values of phenotypic measurements in y ; A_H , with "H" of "High," is defined as the allele that increases the values of phenotypic measurements in y . Alleles are defined in this manner, because Low and High are unique attributes that can be assigned to alleles. In m , a and d are referred to as the additive and dominance effect at the single locus, where a is positive, so that definition of the Low and High attributes is consistent. Actual computations were based on nonuniquely defined alleles A_1 and A_2 with unrestricted a ; uniquely defined alleles A_L and A_H with $a \geq 0$ were obtained as a transformation (see APPENDIX).

Above, distinction between the two heterozygotes is made to allow for a flexible notation of pedigree genotype probabilities as follows: $\Pr(A_e A_f) = p_e p_f$ for Meishan founder animals, $\Pr(A_e A_f) = r_e r_f$ for Dutch founder animals, with $e, f \in \{L, H\}$ and where p_L and p_H ($p_L + p_H = 1$) are the frequency of alleles A_L and A_H in Meishan founders and r_L and r_H ($r_L + r_H = 1$) are the frequency of alleles A_L and A_H in Dutch founders; $\Pr(A_e A_f) = \tau_{e,gh} \tau_{f,g^*h^*}$, for all nonfounder individuals, with $e, f, g, h, g^*, h^* \in \{L, H\}$, and where $A_g A_h$ and $A_{g^*} A_{h^*}$ are the genotypes of the sire and dam of the individual considered, $\tau_{L,gh}$ is the transmission probability for genotype $A_g A_h$ to transmit an A_L allele and $\tau_{H,gh} = 1 - \tau_{L,gh}$ is the corresponding probability to transmit the A_H allele. With Mendelian inheritance $\tau_{L,LL} = 1$, $\tau_{L,LH} = \tau_{L,HL} = 1/2$ and $\tau_{L,HH} = 0$. Distributional assumptions for e are specified as $e \sim N(0, I\sigma_e^2)$ and for u are specified as $u \sim N(0, A\sigma_u^2)$, where A is the numerator of the relationship matrix. Statistical inference was based on a Bayesian approach. Specification of the statistical model for the Bayesian approach is completed by specifying use of uniform prior distributions for nongenetic effects, variance components, effects at the single locus and allele frequencies. These prior distributions were defined on $[-\infty, \infty]$ for the nongenetic effects and effects at the single locus, on $[0, \infty]$ for the variance components and on $[0, 1]$ for the allele frequencies. Variances were assumed *a priori* positive, which was computationally implemented by defining the prior on $[10^{-12}, \infty]$. The prior defined on $[-\infty, \infty]$ for the additive effect at the single locus corresponds to the situation in the actual computations as shortly described above and as exemplified in the APPENDIX. With defined distributional assumptions, the complete set of parameters for model (1) then was $\theta_1 = (\beta, u, W, \sigma_e^2, \sigma_u^2, a, d, p_L, r_L)$. In this set of parameters, the variance components, effects at the single locus and the allele frequencies are referred to as the (genetic) hyperparameters.

As well as a model postulating mixed inheritance, a model postulating pure polygenic inheritance was used by suppressing the term for single gene effects in (1), leading to the model:

$$y = X\beta + Zu + e, \quad (2)$$

with all specifications equal to those of model (1) and with parameters $\theta_2 = (\beta, u, \sigma_e^2, \sigma_u^2)$. The pure polygenic model was used to supply an overall quantification of genetic variance for the traits analyzed. As a third model, a mixed model of inheritance with a restriction on the degree of dominance was used. This model is the same as model (1) except that genotypic values are defined by $m' = (-a, ca, ca, a)$, where c defines the imposed degree of dominance. This model was used to impose complete dominance of the A_L allele ($c =$

–1) or complete dominance of the A_H allele ($c = 1$). In this model also one of the allele frequencies was assumed known, arbitrarily taken to be p_L , such that the set of parameters for this model was $\theta_3 = (\beta, u, W, \sigma_e^2, \sigma_u^2, a, r_L)$.

Gibbs sampling: Bayesian marginal posterior distributions of model parameters were obtained using Gibbs sampling. In such an approach, a Markov chain is constructed that is known to have a stationary distribution equal to the joint posterior distribution of all model parameters, here all parameters in θ_1 for the main model (1), θ_2 for the polygenic model, or θ_3 for the model with restricted degree of dominance. From such a Markov chain samples of marginal posterior distributions of model parameters and of functions of model parameters were obtained. The construction of such a Markov chain was described by JANSSE *et al.* (1995) and implemented in a software package (see APPENDIX). This implementation includes blocked sampling of genotypes of each sire with those of its final progeny and similar blocked sampling of polygenic effects of each sire with those of its final progeny (JANSSE *et al.* 1995). This blocked sampling facilitates convergence of the Gibbs sampler when analyzing typical animal breeding data sets with relatively large progeny groups. Additionally, a model-relaxation technique (SHEEHAN and THOMAS 1993) was applied to further improve convergence of the Gibbs sampler for the single gene component, by relaxation of transmission probabilities. Such a relaxation uses $\tau_{L,LL} = 1 - p_{nmt}$ and $\tau_{L,HH} = p_{nmt}$, where p_{nmt} is a small probability for “non-Mendelian transmission”. Inference about the strict Mendelian model of interest is made by using from the constructed Markov chains only those samples where the genotype configuration was Mendelian. SHEEHAN and THOMAS (1993) showed that the rate at which Mendelian samples randomly appear in a relaxed chain equals the likelihood ratio between the strict Mendelian model and the relaxed model, dependent therefore not only on the parameter p_{nmt} , but also on the data. To achieve a certain rate of Mendelian samples, some trial runs are required to determine, for each data set analyzed, a suitable value of p_{nmt} . For the relaxation technique to have a reasonable impact on convergence, relaxation may be relatively strong, leading to a low rate of Mendelian samples in the relaxed chains: in the analyses performed here, we aimed at a rate of Mendelian samples of 1–10%. Gibbs chains computed for inferences in the mixed inheritance model were started as a “hot” chain (in the terminology of LIN *et al.* 1993), using initially $p_{nmt} = 0.5$, which defines a nongenetic transmission model by allowing random transmission of alleles. Subsequently, such a hot chain was annealed by slowly reducing p_{nmt} to near zero, which restricts movement of the chain to the Mendelian and near-Mendelian space. In the construction of the Gibbs sampler, sampling of random realizations for various types of distributions was based, directly or indirectly, on the uniform random number generator RAN2 (PRESS *et al.* 1992). For construction and sampling details see JANSSE *et al.* (1995) and the APPENDIX.

Convergence of the Gibbs sampler was judged for the hyperparameters by comparison of samples from replicated chains by analysis-of-variance (ANOVA), testing for a significant chain effect. In this approach, Gibbs chains are run that are sufficiently long to obtain a number of independent samples from each chain. This, then, allows to test for equality of the within- and between-chain variances with a standard ANOVA F-test. Significant differences between chains are considered an indication of (practical) reducibility, in which case Gibbs sampling theory (GEMAN and GEMAN 1984; GELFAND and SMITH 1990) does not hold and the samples generated are not from the correct marginal distributions. In this case, the Gibbs sampler is said not to have converged. Significance of differences was assumed when the F-statistic exceeded the

1% significance level. The ANOVA requires independence of the samples, hence only a number of states from each chain, sufficiently spaced, are used. Determination of a suitable spacing yielding virtually serially independent samples was done according to the procedure exemplified by JANSSE *et al.* (1995), mainly based on RAFTERY and LEWIS (1992). The same spacing was used for all parameters of interest. The ANOVA also acts as a post-check on the presumed independence of the Gibbs samples: when the spacing between samples is not sufficient and the assumption of independence does not hold, computed F-statistics will be inflated and chain-effects could be found significant. Insufficient spacing can be verified by increasing the spacing between samples used in the ANOVA by running longer Gibbs chains that keep the same number of samples in the ANOVA to not affect power to detect differences between the chains.

Statistical inference: Generated independent samples used in the ANOVA as described above for assessing convergence are subsequently used for statistical inferences. From samples of marginal posterior distributions, nonparametric density estimates of posteriors were made in the form of average shifted histograms (SCOTT 1992). Such a graph provides a more general and broad inference, than a specific point and/or interval estimator. For parameters with natural boundaries on their parameter space, density estimates were smoothed up to the bound(s) of the parameter space by a reflection boundary technique (SCOTT 1992). Second, samples from marginal posterior distributions were used to compute estimates of mean and standard deviation of the posterior distributions, which were estimated by mean and standard deviation of the Gibbs samples. These estimates converge stochastically, with increasing number of Gibbs samples generated, to the true mean and standard deviation of the marginal posterior distributions of the respective parameters (SMITH and ROBERTS 1993). The posterior mean was chosen used as a point estimator, falling in the class of APE (*a posteriori* expectation) estimators. Such APE estimators have the general property of minimizing quadratic posterior loss. The higher marginalized Bayesian estimators, compared to classical maximum likelihood estimators, are expected to have the same asymptotic properties and superior nonasymptotic properties from a Bayesian viewpoint (GIANOLA and FOULLEY 1990). In analogy to frequentist approaches for statistical inference, the posterior standard deviation can be interpreted as a standard error of the parameter estimate, but is not in general equal to a frequentist standard error.

Of primary interest for statistical inferences were the variance components σ_u^2 for polygenic variance and σ_w^2 for the variance explained by the single gene. These two genetic variances were used to judge significance of the genetic model and, in particular, to judge significance of the single gene component in the model. The variance of single gene effects σ_w^2 was computed as a function of effects at the single locus and of allele frequencies in each Gibbs sample as $2pq(a + d(q - p))^2 + (2pqd)^2$ (FALCONER 1989), where a and d are as defined previously, p is the frequency of the favorable A_H allele, and $q = 1 - p$. The variance of single gene effects was computed to represent the variance in the F_2 generation, by using $p = (p_H + r_H)/2$ in this formula. Nonsignificance of a variance component (shortly σ^2) was empirically shown to lead to a posterior distribution with global mode at $\sigma^2 = 0$ (JANSSE *et al.* 1995). Significance of a variance component shows a global mode for $\sigma^2 > 0$, which may still be accompanied by a local mode or a nonzero density at $\sigma^2 = 0$. For variance of single gene effects, σ_w^2 , a penalty is used to reduce the error of falsely accepting presence of a single gene, by considering σ_w^2 to be significant only when a global mode for $\sigma_w^2 > 0$ has a density 20-fold larger than the density at $\sigma_w^2 =$

0, corresponding to a 5% significance level. In this manner the usual conservatism is applied, accepting presence of a single gene only when abundant evidence is available, or else not rejecting the null hypothesis of polygenic inheritance. The mode(s) and density ratios were determined from the nonparametric density estimates. Once a mixed inheritance model is found likely, further inferences focused on the effects at the single locus and on allele frequencies. Allele frequencies are not uniquely identifiable with the available data on F_2 crossbreds only. For instance, it is not possible to distinguish between a case with $p_L = 1$ and $r_L = 0$ (origin of A_L from Meishan founders) and a case with $p_L = 0$ and $r_L = 1$ (origin of A_L from Dutch founders). The available data only allows unique determination of the genotype frequencies in F_1 parents, yielding estimable functions of allele frequencies being $p_L r_L$, $p_L r_H + p_H r_L$ and $p_H r_H$, representing the frequencies of $A_L A_L$, $A_L A_H$ and $A_H A_L$, and $A_H A_H$ genotypes in the F_1 generation.

From the Gibbs chains, also marginal posterior distributions of individual genotypes, referred to as genotype probabilities, were estimated from the frequency counts of the genotypes sampled in the Markov chain. Genotype probabilities could be estimated from Gibbs chains sampling (β, u, W) conditional on some point estimates for the hyperparameters. Here, however, genotype probabilities were estimated from Gibbs chains sampling all model parameters, which will supply estimates of genotype probabilities not conditional on any point estimates for hyperparameters and where uncertainty from estimation of hyperparameters will be included. Estimates of genotype probabilities were used to study whether various traits found to be influenced by a single gene, actually could be influenced by the same gene, as follows: the interval $[0,1]$ in which genotype probabilities fall, was discretized into $k = 1, \dots, K$ smaller intervals; for a group of n individuals, the number of individuals was counted with a genotype probability for a first trait falling in interval k and with a genotype probability for a second trait falling in interval k^* , the count being denoted c_{kk^*} , for $k = 1, \dots, K$ and $k^* = 1, \dots, K$; the quantities c_{kk^*} were collected in a K -by- K table in which association between the genotypes was tested by a chi-squared test for association with $(K-1)(K-1)$ degrees of freedom; a significant association was considered to be an indication that two traits, found to be influenced by single genes, could actually well be influenced by the same gene affecting both traits. This procedure works best with large K , but the choice for K is bounded by the number of individuals n because counts c_{kk^*} must be reasonably large for the approximate chi-squared test to be valid. Choice of the interval cut-points is arbitrary and also can be chosen such that all counts c_{kk^*} are reasonably large. Estimated genotype probabilities preferably should have high and similar accuracy.

RESULTS

Polygenic model: For inference in the polygenic model, for each trait a trial Gibbs chain of 10,000 cycles was run. From these chains, it was determined that virtually independent samples could be obtained using a spacing of 800 cycles. Subsequently, for inferences, for each trait five Gibbs chains of 40,000 cycles were run, obtaining 50 independent Gibbs samples per chain and 250 samples in total per trait. Using an estimate of phenotypic variance from a model fitting nongenetic effects only, starting values for polygenic variances in the five replicated chains were chosen as 10, 20, 30, 40 and 50% of this estimate of phenotypic variance, and with error

variance equal to the remainder. Starting values for nongenetic effects and polygenic effects were zero. A burn-in period of 1600 cycles was used to allow the Gibbs chains to reach equilibrium. Estimated posterior means and posterior standard deviations for variance components and heritability are in Table 2. Features of the posterior distribution for heritability (h^2) were obtained by computing from each Gibbs sample of variance components the corresponding value of h^2 and subsequently using these values to summarize the posterior distribution of h^2 . Tests for convergence of the Gibbs sampler by comparison of multiple chain output using ANOVA on the independent samples showed no significant differences between replicated chains for all parameters in Table 2, demonstrating convergence of the Gibbs sampler. Results indicate general existence of genetic variation for the various traits. Absolute values of heritability are generally low: HOVENIER *et al.* (1993) indicate in a review average heritabilities for water-holding capacity traits (Cook, Drip) and pH of 0.20, for Shear and color traits of 0.30, and for Imfat of 0.50. Heritabilities in the data analyzed are 0.10 to 0.20 below these average literature values.

Mixed inheritance model: For the mixed inheritance model, trial Gibbs chains were run to determine suitable values of the relaxation parameter and the spacing to be used between Gibbs cycles to yield independent Gibbs samples. Such trial runs showed that for some traits too strong relaxation could lead to Gibbs chains settling at non-Mendelian states, without further realizations of Mendelian samples. Therefore, a variety of approaches was used to obtain Mendelian samples for the mixed inheritance model: for "well behaved" traits, which were Cook, Imfat, Fat, Lean, White and Yellow, a relatively large relaxation was used leading to a low rate of Mendelian samples of 1–2% and virtually independent Mendelian samples were obtained by taking every 20th Mendelian sample occurring; for the remaining "less behaved" traits, a smaller relaxation was used leading to a larger rate of Mendelian samples of 5–10% and virtually independent Mendelian samples were obtained by taking, varying between traits, every 50th to 180th Mendelian sample occurring. In this manner, one Mendelian sample was obtained about every 1000–2000 cycles for all traits analyzed. Values for the relaxation parameter p_{nmt} to obtain the mentioned rates of non-Mendelian samples ranged from 2.0×10^{-3} to 6.5×10^{-3} . For inferences, five chains per trait were run, obtaining 50 independent Mendelian samples per chain and 250 in total per trait. All chains started with the described annealing of a hot chain, which was taken over 1000 cycles, and was followed by a burn-in of another 1000 cycles. Gibbs chains were started using the following parameter values: zeros for nongenetic effects, polygenic effects and effects at the single locus, heterozygotes for all genotypes, posterior mean estimates from the polygenic model (Table 2) for variance

TABLE 2
Estimated variance components in the polygenic model

Trait	σ_e^2		σ_u^2		h^2	
	mpm	mpsd	mpm	mpsd	mpm	mpsd
Cook	7.31	0.503	1.07	0.578	0.126	0.064
Drip	1.71	0.134	0.272	0.127	0.136	0.061
Shear	51.2	2.93	2.45	2.07	0.045	0.038
Imfat	0.429	0.0544	0.258	0.0757	0.372	0.094
pH ^a	4.39	0.348	0.567	0.313	0.114	0.061
pH-s ^a	5.06	0.385	0.551	0.354	0.098	0.060
Light	15.8	1.35	3.61	1.47	0.184	0.071
Red	2.63	0.200	0.431	0.198	0.140	0.061
Yellow	2.60	0.190	0.294	0.163	0.101	0.054
Fat	19.4	1.79	6.26	2.36	0.241	0.081
Lean	33.4	2.76	8.08	3.25	0.193	0.073

Marginal posterior means (mpm) and marginal posterior standard deviations (mpsd) for error variance (σ_e^2), polygenic variance (σ_u^2) and heritability (h^2), based on a total of 250 independent Gibbs samples from five replicated chains.

^a For pH and pH-s variance components are in hundreds.

components and 0.5 for allele frequencies. It was observed that the approach of annealing a hot chain was quite effective in letting the Gibbs sampler converge to its equilibrium, even from such a crude starting point.

Estimated marginal posterior distributions for the two genetic variance components in the mixed inheritance model are in Figure 2 and estimated marginal posterior means and standard deviations of all three variance components are in Table 3. Analysis of differences between replicated chains indicated two traits where lack of convergence of the Gibbs sampler was diagnosed: pH-s and Fat, for parameters residual variance and polygenic variance. Convergence was found for estimation of single-gene variances for all traits. Based on the marginal posterior distributions depicted in Figure 2, traits were grouped according to significance of the two genetic variances as follows: (1) traits Cook, Drip and Imfat showing significant influence of a single gene in presence of additional significant polygenic variance, (2) traits Shear, pH, pH-s, Fat and Lean, showing significant influence of a single gene, but with low polygenic variances showing nonnegligible densities at $\sigma_u^2 = 0$, (3) color traits Light, Red and Yellow, showing nonsignificant single gene variance. Group 2 includes the two traits with no convergence of the Gibbs sampler for some parameters. The total of genetic variance inferred in the mixed inheritance model appears larger than genetic variance inferred in the pure polygenic model. This can be explained by the segregation variance at the single locus, which will be attributed to error variance in the polygenic model, but can be included in genetic variance in the mixed inheritance model.

The eight traits in groups 1 and 2 described above were considered for further investigations on estimates of effects at the single locus (Table 4) and F_1 genotype

frequencies (Table 5). Estimates for additive effect a and dominance effect d at the single locus indicated that d was likely to be of the same absolute value than a : Table 4 shows the estimated 95% highest posterior density (HPD) regions for the difference $a - |d|$, which in all cases included the value zero. HPD regions were obtained by computing, from each Gibbs sample of a and d , the difference $a - |d|$, subsequently making a nonparametric density estimate for this difference and obtaining from this density estimate the left and right 2.5% quantiles. Hence, a single gene with complete dominance for one of its alleles was inferred for all traits listed in Table 4. Reservations on this conclusion should be made for pH-s and Fat, where nonconvergence of the Gibbs sampler was diagnosed for estimation of the dominance effect at the single locus. Estimates of genotype frequencies in the F_1 (Table 5) indicated absence of the homozygote recessive genotype in the F_1 parents. The frequency of homozygote recessives in F_1 is $p_L r_L$ for Cook (with d positive) and $p_H r_H$ for other traits (with d negative). Posterior distributions for $p_L r_L$ for Cook and $p_H r_H$ for other traits are in Figure 3 for the eight traits considered and show global modes at zero for all these frequencies. Absence of the homozygote recessive genotype in F_1 indicates that the recessive allele must be absent in one of the founder lines, although the observations on F_2 used here, do not allow one to determine the founder line. Convergence of the Gibbs sampler for estimation of the genotype frequency of the homozygote recessives was confirmed for all traits.

Restricted model: Results presented in Tables 4 and 5 led to the conclusion that a number of traits might be influenced by a recessive gene, absent in one of the founder lines. However, for traits pH-s and Fat, convergence of the Gibbs sampler was not observed,

FIGURE 2.—Estimated marginal posterior distributions (averaged histogram frequencies) for polygenic variance (σ_u^2) and single gene variance (σ_w^2) in the mixed inheritance model for 11 meat quality traits.

TABLE 4
Estimated effects at the single locus in the mixed inheritance model

Trait	<i>a</i>		<i>d</i>		<i>a</i> - <i> d </i> 95% HPD region	
	mpm	mpsd	mpm	mpsd	From	To
Cook	4.67	0.436	4.64	0.562	-1.08	1.29
Drip	1.40	0.0886	-1.53	0.175	-0.631	0.323
Shear	5.72	0.844	-8.54	1.74	-5.99	1.21
Imfat	1.14	0.0767	-1.09	0.128	-0.239	0.359
pH	0.319	0.0170	-0.313	0.0241	-0.0475	0.0651
pH-s	0.233	0.0202	-0.269**	0.0493	-0.139	0.0831
Fat	4.39	0.584	-3.85**	1.07	-2.32	3.38
Lean	4.39	1.05	-4.16	1.67	-4.83	4.67

Marginal posterior means (mpm) and marginal posterior standard deviations (mpsd) for additive effect (*a*) and dominance effect (*d*) at the single locus, and estimated 95% highest posterior density (HPD) regions for their difference, based on a total of 250 independent Gibbs samples from five replicated chains. Estimates are shown for traits with significant contributions of single gene variance. ** Significant differences between replicated chains ($P < 0.01$).

which did not allow one to draw definite conclusions. Also, the finding of several traits being influenced by a single gene brings up the interesting hypothesis of whether all or some traits might actually be influenced by the same gene, acting pleiotropically. By using a restricted model, which assumed complete dominance and absence of the recessive allele in one of the founder lines, it was attempted to improve inference for pH-s and Fat, and to obtain a more accurate estimation of genotype probabilities compared to an unrestricted model. It was not considered to obtain genotype probabilities conditional on some point estimates for all hyperparameters because we thought that such an approach could endanger attempts to elucidate pleiotropic effects of the genes.

Restrictions to impose complete dominance and absence of the recessive allele in one of the founder lines

were applied as follows (*cf.* Tables 4 and 5): for Cook, $d = a$, $p_L = 0$ and $p_H = 1$; for all other traits, $d = -a$, $p_L = 1$ and $p_H = 0$. Application of the restriction to allele frequency p_L is arbitrary from a modelling viewpoint, as only the genotype frequencies in F_1 are uniquely estimable, but was thought to be better from a computing viewpoint: fixing p_L , frequency in the (smaller) paternal founder line, and estimating p_L , frequency in the (larger) maternal founder line was thought to supply a more stable Gibbs chain. Gibbs chains were started using the following parameter values: zeros for nongenetic effects and polygenic effects, heterozygotes for all genotypes, posterior mean estimates in the mixed inheritance model (Tables 3 and 4) for variance components and additive effect at the single locus *a*, and 0.5 for allele frequency p_L . For inference in this restricted model, an initial phase where a hot chain was annealed

TABLE 3
Estimated variance components in the mixed inheritance model

Trait	σ_e^2		σ_u^2		σ_w^2	
	mpm	mpsd	mpm	mpsd	mpm	mpsd
Cook	5.03	0.417	0.931	0.399	3.73	1.30
Drip	0.582	0.0697	0.142	0.0627	1.29	0.158
Shear	35.2	3.32	2.55	1.98	18.7	4.79
Imfat	0.260	0.0310	0.120	0.0416	0.351	0.0914
pH ^a	2.03	0.165	0.220	0.156	3.28	0.663
pH-s ^a	2.72**	0.377	0.344**	0.216	2.85	0.623
Light	12.9	2.61	3.34	1.42	3.59	2.46
Red	2.29	0.304	0.302	0.209	0.501	0.314
Yellow	2.29	0.274	0.233	0.156	0.436	0.797
Fat	12.6**	1.82	3.34**	1.94	9.92	2.24
Lean	21.7	3.97	5.01	3.19	14.4	3.81

Marginal posterior means (mpm) and marginal posterior standard deviations (mpsd) for error variance (σ_e^2), polygenic variance (σ_u^2) and single gene variance (σ_w^2), based on a total of 250 independent Gibbs samples from five replicated chains. ** Significant differences between replicated chains ($P < 0.01$).

^a For pH and pH-s variance components are in hundreds.

FIGURE 3.—Estimated marginal posterior distributions (averaged histogram frequencies) for the frequency of the double recessive genotype in F_1 parents for eight meat quality traits: vertical positions of trait names indicated starts of the graphs at the left boundary for the respective traits.

was omitted: starting with plausible values for effects at the single locus, the Gibbs sampler converged equally well without such phase. Other details were the same as for the full mixed inheritance model, generating a total of 250 Gibbs samples per trait from five replicated chains.

Estimates of variance components with the restricted model confirmed estimates for the full model (not shown). Estimates for the effect at the single locus and frequency of the favorable A_H allele are in Table 6. Due to the restriction in the genetic components of the model, estimates of genetic variances were slightly lower, and estimates of error variance were slightly

higher. The Gibbs sampler showed good convergence for the estimation of variance components for all traits, except for the estimation of residual variance for Lean, and results showed significant contributions of single gene variance for all traits. For Lean, influence of a single gene was rejected, first because of poor convergence of the Gibbs sampler, and second, because for Lean frequency of the recessive allele, which was forced to zero in one of the founder lines, was estimated very close to one in the other founder line (Table 6). In such case, inferred significant single gene variance could be caused by a general nonnormality in the data (see DISCUSSION). For the seven remaining traits, Cook, Drip, Shear, Imfat, pH, pH-s and Fat, influence of a single gene is considered very likely.

From the Gibbs chains used to estimate genetic parameters in the reduced model (partly in Table 6), estimates of genotype probabilities were obtained as well. Using genotype probabilities inferred for different traits, association tests were carried out as described in MATERIALS AND METHODS to obtain indications of whether traits presumed to be affected by a single gene could actually be affected by the same gene. Associations were studied between seven traits, which were all traits in Table 6 except Lean. Tests for associations were based on estimated probabilities of female F_1 parents to be heterozygote. Use of the female parents supplied a reasonable compromise between requiring a large number of individuals and requiring individuals with precise estimates. Using female parents, 251 individuals were available and the genotype probabilities were counted in three intervals: those falling between 0 and 0.65, between 0.65 and 0.80 and between 0.80 and 1. These intervals supplied a good distribution of numbers of probabilities falling in all combinations of intervals for pairs of traits and resulted in an appropriate use of the chi-squared approximation to test for association

TABLE 5

Estimated F_1 genotype frequencies in the mixed inheritance model

Trait	$p_L r_L$		$p_L r_H + p_H r_L$		$p_H r_H$	
	mpm	mpsd	mpm	mpsd	mpm	mpsd
Cook	0.0189 ^a	0.0166 ^a	0.379	0.0631	0.602	0.0683
Drip	0.161	0.0575	0.806	0.0657	0.0325 ^a	0.0309 ^a
Shear	0.381	0.105	0.568	0.106	0.0515 ^a	0.0419 ^a
Imfat	0.495	0.0666	0.471	0.0639	0.0334 ^a	0.0265 ^a
pH	0.427	0.0601	0.551	0.0606	0.0218 ^a	0.0208 ^a
pH-s	0.294**	0.111	0.673**	0.115	0.0337 ^a	0.0305 ^a
Fat	0.254	0.101	0.702	0.108	0.0431 ^a	0.0409 ^a
Lean	0.129	0.0905	0.778	0.113	0.0937 ^a	0.0749 ^a

Marginal posterior means (mpm) and marginal posterior standard deviations (mpsd) for the genotype frequencies in F_1 , based on a total of 250 independent Gibbs samples from five replicated chains. Genotype frequencies in F_1 are the estimable functions of the allele frequencies p_L and r_L . Estimates are shown for traits with significant contributions of single gene variance. ** Significant differences between replicated chains ($P < 0.01$).

^a Global mode at zero.

TABLE 6

Estimated single-gene parameters in the restricted mixed inheritance model

Trait	a		η_H	
	mpm	mpsd	mpm	mpsd
Cook	4.64	0.368	0.614	0.0652
Drip	1.44	0.0460	0.842	0.0590
Imfat	1.12	0.0665	0.504	0.0580
Shear	6.29	0.720	0.676	0.1000
pH ^a	0.316	0.0153	0.565	0.0606
pH-s ^a	0.240	0.0154	0.734	0.0905
Fat	4.18	0.365	0.782	0.0852
Lean	4.32	0.475	0.91** ^b	0.0752 ^b

Marginal posterior means (mpm) and marginal posterior standard deviations (mpsd) for additive effect at the single locus (a) and allele frequency for allele A_H in Western founders (η_H) based on a total of 250 independent Gibbs samples from five replicated chains. The restricted model assumed complete dominance, with allele A_L the recessive allele for Cook ($d = a$) and with allele A_H the recessive allele for other traits ($d = -a$), and assumed absence of the recessive allele in one of the founder lines, arbitrarily taken as the Meishan founders. Estimates are shown for traits with significant contributions of single gene variance. ** Significant differences between replicated chains ($P < 0.01$).

^a For pH and pH-s variance components are in hundreds.

^b Global mode at $\eta_H = 1$.

for all combinations of traits except one. Using a discretization into three intervals, tests for associations were based on a 3-by-3 table and test statistics followed a chi-squared distribution with four degrees of freedom under the null-hypothesis. Significant associations were considered as those where the test statistic exceeded the 1% level of significance.

Test statistics for associations between genotypes for the seven traits considered are in Table 7. Traits in this table are ordered corresponding to a suggested division into two groups: Cook, pH, pH-s and Fat as a first group and Imfat, Shear and Drip as a second group. Test statistics for association of genotypes between pairs of traits across these groups are nonsignificant, whereas associations between pairs of traits within groups often are significant. In the first group, clear associations were found between combinations of Cook, pH and pH-s, which strongly suggest that these three traits are influenced by the same gene. Also Fat may be influenced by this Cook/pH gene, but here the situation is not clear: Fat was associated with pH-s, but not with any of the other traits in this group. In the second group, the situation also is not fully clear: here Imfat is associated with Shear, Shear with Drip, but Imfat is not associated with Drip.

DISCUSSION

Validation: To argue for presence of a single gene affecting the traits considered, one must demonstrate

TABLE 7

Chi-squared test statistics for association between inferred genotypes of individuals for single genes affecting the different traits

Trait	pH	pH-s	Fat	Imfat	Shear	Drip
Cook	57.7**	24.1**	6.14	3.20	5.83	6.72
pH	—	34.7**	5.44	10.1 ^a	1.94	6.61
pH-s		—	13.6**	5.09	5.05	5.14
Fat			—	1.98	6.91	3.20
Imfat				—	20.7**	3.83
Shear					—	21.3**

Probabilities (P_c) were estimated for 251 female F_1 parents to carry the recessive allele. Estimation was done from Gibbs chains for inference in the restricted model assuming complete dominance and absence of the recessive allele in one of the founder lines. Test for association was based on transformation of estimated probabilities to a three-class variable, indicating whether $P_c < 0.65$, $0.65 < P_c < 0.80$, or $P_c > 0.80$. ** Significant association ($P < 0.01$).

^a Chi-squared approximation not good.

the presence of typical data characteristics and argue that these characteristics are due to segregation of a single gene, rather than some other mechanism. The data characteristics that are typical for traits influenced by a single gene are heterogeneous within family variation for the traits measured and general, or family-specific skewness and/or kurtosis in the distribution of the trait. These characteristics, when they can be observed in certain families but not in others, are fairly robust identifiers for presence of a single gene. However, non-family-specific data characteristics such as general skewness or kurtosis are much less robust identifiers (LE ROY and ELSEN 1992). For the trait Lean, the recessive allele was inferred to be fixed in the founder line where it originated from. This implies that all F_1 parents were inferred to be heterozygous and, apparently, the trait did not show differences in within-family variation or family-specific skewnesses. This led us to reject influence of a single gene on this trait: the general skewness could be the result of a segregating single gene, but not necessarily, so that convincing evidence is not supplied.

For the remaining traits, Cook, pH, pH-s, Fat, Imfat, Shear and Drip, the recessive allele was inferred to segregate in the founder line where it originated from. This implies presence of two genotypes in F_1 parents: parents that carry the recessive allele and parents that do not carry the recessive allele. When mating various combinations of such parents, heterogeneous within-family variation and family-specific skewnesses should arise. Such effects were indeed present for the traits analyzed (Table 8): families of carrier fathers showed markedly higher variance in the measured traits than families of noncarrier fathers. This heterogeneity in within family variances is not easily explained by effects other than the segregation of a single gene because design and analysis of the experiment excluded con-

TABLE 8
Family means and variances for different carrier-status of their fathers

Trait	Noncarrier father families		Dubious father families		Carrier father families	
	N		N		N	
Cook	322	26.5 ± 2.87	205	26.4 ± 3.12	318	26.2 ± 4.15
Drip	0	—	97	2.04 ± 0.943	747	2.78 ± 1.58
Imfat	227	1.65 ± 0.654	260	1.75 ± 0.704	344	2.02 ± 1.05
Shear	62	34.8 ± 6.89	408	36.5 ± 8.51	375	43.7 ± 11.4
pH	162	5.60 ± 0.161	107	5.65 ± 0.229	576	5.68 ± 0.289
pH-s	47	5.71 ± 0.137	149	5.77 ± 0.240	650	5.84 ± 0.322
Fat	26	19.9 ± 4.48	225	20.4 ± 4.76	595	22.7 ± 5.92

N_i number of observations; values are means ± SD of raw phenotypic measurements for traits inferred to be influenced by single genes, in families of non-carrier fathers, families with dubious status of the father and in families of carrier fathers. Carrier-status was based on estimated probabilities (P_c) for F_1 fathers to carry the recessive allele. Estimation was done from Gibbs chains for inference in the restricted model assuming complete dominance and absence of the recessive allele in one of the founder lines. Fathers with $P_c < 0.20$ were considered noncarriers, fathers with $P_c > 0.80$ were considered carriers, and those remaining were considered "dubious". For Drip, none of the fathers was found carrier.

founding with any common factors known to cause possible heterogeneity in variance. In particular, animals were raised at the same time in all locations/companies, which were geographically not widely spread, slaughtered in one slaughter house and a confounding between fathers/families and locations/companies was eliminated by formation of the described central pool of fathers, so that each resulting father-family was a mix of individuals from different locations/companies. This all supplies sufficient evidence in favor of the heterogeneity in family variances for Cook, pH, pH-s, Fat, Imfat, Shear and Drip to be of genetic origin, caused by segregation of a single gene.

Single genes: By simple association tests between inferred genotypes, indications were obtained whether (groups of) traits actually could be influenced by the same gene. We postulated as a working hypothesis that the effects observed could be caused by two genes: one gene that influences cooking loss, pH and possibly backfat thickness, and a second gene that influences intramuscular fat, shearforce and possibly drip loss. The presumed first gene is called Meishan Cooking loss gene (*MC*), the second gene is called Meishan Intramuscular fat gene (*MI*). A joint effect of the *MC* gene on cooking loss and pH is physiologically well understandable and estimated effects of the inferred recessive allele to decrease cooking loss and increase pH agree with expectations from a physiological viewpoint. Whether the *MC* gene also influences backfat thickness remains unclear: an association was found between backfat and one of the pH measures, but not with cooking loss and a second pH measure, and physiologically such an association is also not immediately obvious. For the presumed *MI* gene, the situation is less clear: from a physiological viewpoint, higher intramuscular fat could be associated with lower shearforce, indicative of more tender meat, but from the analyses, higher intramuscular fat appeared associated with higher shearforce. For the *MI*

gene, a possible relationship with drip loss is also debatable, as such an association was only made via shearforce. Therefore, the working hypothesis of only two genes, *MC* and *MI*, influencing the traits analyzed could well be too restrictive and could require extension, postulating effects of more genes.

The recessive alleles of the inferred single genes were found to originate from one of the founder lines only. This raises the interesting question whether this was the Chinese Meishan founder line or the collection of Western founder lines. In the data, some additional evidence for one or the other hypothesis was available. In the analyses, Western founder animals were treated as a homogeneous group, but in fact these founders consisted of different lines, one from each company (Figure 1). Among F_1 fathers that were carriers of the recessive allele for various traits, descendants from all Western founder lines were present. Using this additional information, it is unlikely that such recessive alleles would have been present in all these Western lines, and a more plausible explanation is that these recessive alleles originated from the common Meishan fathers of F_1 fathers.

The above described *MC* gene, affecting pH, superficially might be presumed to be actually the known Halothane gene (EIKELNBOOM and MINKEMA 1974), or the *RN* gene (LE ROY *et al.* 1990), either of which also affects pH in meat. However, presence of the mutation causing Halothane susceptibility was excluded by molecular typing and presence of the *RN*⁻ allele of the *RN* gene is unlikely because this allele is thought to be specific for the Hampshire pig breed. Moreover, effects of the presumed *MC* gene are opposite to those known for the mutation of the Halothane gene and for the *RN*⁻ allele: these two alleles are (partly) dominant and increase cooking loss and decrease pH, whereas the dominant allele of the *MC* gene decreases cooking loss and increases pH.

Bayesian segregation analysis: The secondary goal of this study was to apply a recently developed Bayesian approach for segregation analysis in extensive field data analyses. The approach uses Gibbs sampling for computing marginal posterior distributions and appeared generally feasible. The approach was effective in generating a reasonable number of independent samples from the marginal posterior distributions of parameters, and convergence was found, at least for the variance components, in practically all cases. Gibbs sampling allows use of looped pedigrees, incorporation of many relationships, and circumvents the intrinsic problems (e.g., HASSTEDT 1982) in marginalizing a joint distribution with respect to both discrete parameters (genotypes) and continuous parameters (polygenic effects and others). Because of these advantages, Gibbs sampling can, and has, also been used in maximum likelihood approaches to segregation analysis (e.g., GUO and THOMPSON 1992). In combination with the Bayesian approach, Gibbs sampling provides even more flexibility than a maximum likelihood approach, for instance in the estimation of means. Models used in analysis of livestock data often comprise a large number of means; in the analysis presented here 33 means and one regression coefficient were fitted. In a Bayesian approach, these means are straightforwardly included in the Gibbs chains and treated as nuisance parameters, yielding a REML-type approach by accounting for uncertainty originating from the estimation of these fixed effects. Accounting for such uncertainty is known, from linear model applications, to remove bias in estimation of variance components. Apart from flexibility in the model, the Bayesian approach supplies posterior distributions and small-sample "standard errors" of parameters, while the maximum likelihood approach relies on asymptotic properties.

Conclusions: The primary aim of this study was to investigate whether meat quality traits in a crossed pig population were influenced by single genes. The statistical analyses presented showed convincingly that seven meat quality traits measured in this population are indeed influenced by single genes, which most likely originate from the Chinese Meishan breed. These genes are different from genes so far identified to affect meat quality and further study of this population will be worthwhile. Currently the animals are being genotyped for a large number of genetic markers that will enable a linkage analysis to estimate location of the genes. Based on the results of the present study it can be concluded that the material from the F_2 cross is very suited to locate genes affecting meat quality. The results of the linkage analysis will, in particular, be helpful to determine the number of genes that are actually responsible for the observed effects on the seven traits analyzed. As a working hypothesis we postulated presence of two genes, called *MC* and *MI*, but the analysis based on phenotypic measurements only leaves considerable uncertainty about this point.

Dutch breeding companies participating in the described crossing experiment were NVS, Bovar, Euribrid, Fomeva and Nieuw-Dalland. Meishan founders used in the crossing experiment are from a pure Meishan herd at Wageningen Agricultural University, made available by Euribrid (Boxmeer, The Netherlands). Research was supported financially by the Dutch Product Board for Livestock, Meat and Eggs, and by the aforementioned breeding companies participating in the experiment.

LITERATURE CITED

- CAMERON, N. D., 1990 Genetic and phenotypic parameters for carcass traits, meat and eating quality in pigs. *Livest. Prod. Sci.* **26**: 119–135.
- EIKELENBOOM, G., and D. MINKEMA, 1974 Prediction of pale, soft, exudative muscle with a non-lethal test for the halothane-induced porcine malignant hyperthermia syndrome. *Tijdschr. Diergeneeskunde* **99**: 421–426.
- ELSTON, R. C., and J. STEWART, 1971 A general model for the genetic analysis of pedigree data. *Hum. Hered.* **21**: 523–542.
- FALCONER, D. S., 1989 *Introduction to Quantitative Genetics*, Ed. 3. Longman, Harlow, London.
- GELFAND, A. E., and A. F. M. SMITH, 1990 Sampling based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**: 398–409.
- GEMAN, S., and D. GEMAN, 1984 Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern. Anal. Mach. Intell.* **6**: 721–741.
- GANOLA, D., and J. L. FOULLEY, 1990 Variance estimation from integrated likelihoods (VEIL). *Genet. Sel. Evol.* **22**: 403–417.
- GUO, S. W., and A. E. THOMPSON, 1991 Monte carlo estimation of variance component models for large complex pedigrees. *IMA J. Math. Appl. Med. Biol.* **8**: 171–189.
- GUO, S. W., and A. E. THOMPSON, 1992 A monte carlo method for combined segregation and linkage analysis. *Am. J. Hum. Genet.* **51**: 1111–1126.
- HASSTEDT, S. J., 1982 A mixed-model likelihood approximation on large pedigrees. *Computer Biomed. Res.* **15**: 295–307.
- HILL, W. G., and S. A. KNOTT, 1990 Identification of genes with large effects, pp. 477–495 in *Advances in Statistical Methods for Genetic Improvement of Livestock*, edited by D. GIANOLA and K. HAMMOND. Springer-Verlag, Berlin.
- HOVENIER, R., E. KANIS, T. VAN ASSELDONK and N. G. WESTERINK, 1992 Genetic parameters of pig meat quality traits in a Halothane negative population. *Livest. Prod. Sci.* **32**: 309–321.
- HOVENIER, R., E. KANIS, T. VAN ASSELDONK and N. G. WESTERINK, 1993 Breeding for meat quality in Halothane negative populations—a review. *Pig News Inform.* **14**: 17N–25N.
- JANSS, L. L. G., R. THOMPSON and J. A. M. VAN ARENDONK, 1995 Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations. *Theor. Appl. Genet.* **91**: 1137–1147.
- KNOTT, S. A., C. S. HALEY and R. THOMPSON, 1992 Methods of segregation analysis for animal breeding data: a comparison of power. *Heredity* **68**: 299–311.
- LANDE, R., 1981 The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* **99**: 541–553.
- LE ROY, P., and J. M. ELSSEN, 1992 Simple test statistics for major gene detection: a numerical comparison. *Theor. Appl. Genet.* **83**: 635–644.
- LE ROY, P., J. M. ELSSEN and S. A. KNOTT, 1989 Comparison of four statistical methods for detection of a major gene in a progeny test design. *Genet. Sel. Evol.* **21**: 341–357.
- LE ROY, P., J. NAVEAU, J. M. ELSSEN and P. SELLIER, 1990 Evidence for a new major gene influencing meat quality in pigs. *Genet. Res.* **55**: 33–40.
- LIN, S., E. THOMPSON and E. WIJSMAN, 1993 Achieving irreducibility if the Markov chain Monte Carlo method applied to pedigree data. *IMA J. Math. Appl. Med. Biol.* **10**: 1–7.
- MACDOUGALL, D. B., 1986 The chemistry of color and appearance. *Food Chem.* **21**: 283–299.
- MORTON, N. E., and C. J. MACLEAN, 1974 Analysis of family resemblance III. Complex segregation of quantitative traits. *Am. J. Hum. Genet.* **26**: 489–503.

- OTSU, K., M. S. PHILLIPS, V. K. KHANNA, S. DE LEON and D. H. MACLENNAN, 1992 Refinement of diagnostic assays for a probable causal mutation for porcine and human malignant hyperthermia. *Genomics* **13**: 835–837.
- PRESS, W. H., B. P. FLANNERY, S. A. TEUKOLSKY and W. T. VETTERLING, 1992 *Numerical Recipes: The Art of Scientific Computing*, Ed. 2. Cambridge University Press, Cambridge, MA.
- RAFTERY, A. E., and S. M. LEWIS, 1992 How many iterates in the Gibbs sampler?, pp. 765–776 in *Bayesian Statistics IV*, edited by J. M. BERNARDO, J. O. BERGEN, A. P. DAVID and A. F. M. SMITH. Oxford University Press, Oxford.
- SAS INSTITUTE INC., 1988 *SAS/STAT Users Guide*, release 6.06. Cary, NC.
- SCOTT, D. W., 1992 *Multivariate Density Estimation*. Wiley and Sons, New York.
- SHEEHAN, N., and A. THOMAS, 1993 On the irreducibility of a Markov chain defined on a space of genotype configurations by a sampling scheme. *Biometrics* **49**: 163–175.
- SMITH, A. F. M., and G. O. ROBERTS, 1993 Bayesian computation via the Gibbs sampler and related markov chain monte carlo methods. *J. R. Stat. Soc. B* **55**: 3–24.
- SORENSEN, D. A., C. S. WANG, J. JENSEN and D. GIANOLA, 1994 Bayesian analysis of genetic change due to selection using Gibbs sampling. *Genet. Sel. Evol.* **26**: 333–360.
- SORENSEN, D. A., S. ANDERSEN, D. GIANOLA and I. KORSGAARD, 1995 Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.* **27**: 229–249.
- TANNER, M. A., 1993 *Tools for Statistical Inference*. Springer, Berlin.
- THOMAS, D. C., and V. CORTESIS, 1992 A Gibbs sampling approach to linkage analysis. *Hum. Hered.* **42**: 63–76.
- WANG, C. S., J. J. RUTLEDGE and D. GIANOLA, 1993 Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genet. Sel. Evol.* **25**: 41–62.
- ZHANG, Y., R. PROENCA, H. MAFFEI, M. BARONE, L. LEOPOLD *et al.*, 1994 Positional cloning of the mouse *obese* gene and its human homologue. *Nature* **372**: 425–432.

Communicating editor: B. S. WEIR

APPENDIX: DETAILS ON CONSTRUCTION OF GIBBS SAMPLERS FOR INFERENCE ON MIXED INHERITANCE MODELS

A software package was developed to construct Gibbs samplers for inference in the mixed inheritance model (1) described in MATERIALS AND METHODS. The main theory on the construction of the required Gibbs sampler is described by JANSS *et al.* (1995). Hereafter, some general information on the developed package is given and extensions to JANSS *et al.* (1995) used in the present study and some computational remarks are described.

General: The set-up of the package is largely a “help-yourself tool-kit,” consisting of a set of FORTRAN-variables, corresponding to model-parameters, and a set of FORTRAN-77 routines to sample (groups of) parameters. By repeatedly calling these routines, Gibbs samplers are generated. Additional routines are supplied that read, order and code pedigree and data files and that make information from these files available to the routines for sampling model parameters. Set-up of the package also allows fitting of submodels of the mixed inheritance model, *e.g.*, the pure polygenic model, and allows suppressing the sampling of parameters, in which case parameters are updated by their “current” expectation. As described by JANSS *et al.* (1995), the latter allows application of a hierarchy of inferential approaches, for example a Monte Carlo

EM likelihood-approach or Gauss-Seidel schemes for iteratively solving of linear model equations can be specified. The software package can be obtained from the authors.

Sampling details: The Gibbs sampler constructed by JANSS *et al.* (1995) included the following: sampling of levels of one or more fixed categorical nongenetic effects, sampling of polygenic effects assuming each individual in the pedigree has one observation and applying blocking of effects of each sire and those of its final progeny, sampling of genotypes in similar blocks as polygenic effects, sampling of variance components using flat priors, sampling of an additive effect at the single locus, and sampling of allele frequency in the (one) founder population. The software package developed contains routines to sample these parameters. For inference in the mixed inheritance models used in the analyses presented here, the following features were added:

- Covariates were allowed for by allowing the design matrix $\mathbf{X}$ for a nongenetic effect to be a single column vector, containing measured covariates;
- Modelling of a dominance effect and modelling of additive and dominance effect with restricted relative dominance effect at the single locus were allowed for. The equation to sample the additive effect a at the single locus given by JANSS *et al.* (1995) can be shown to be based on a linear model to regress the corrected data on a dummy vector $\mathbf{W}'\mathbf{Z}'(-1, 0, 0, 1)$. By analogy, sampling of the dominance effect d at the single locus uses the dummy vector $\mathbf{W}'\mathbf{Z}'(0, 1, 1, 0)$; and sampling of the additive effect at the single locus, assuming $d = ca$, uses the dummy vector $\mathbf{W}'\mathbf{Z}'(-1, c, c, 1)$. In the model with restricted dominance effect, after sampling of a new a , d is set to ca .
- Missing observations were allowed for. For polygenic effects and effects at the single locus this follows straightforwardly from linear model methodology by allowing the $\mathbf{Z}$ matrix to contain columns with all zeroes. For the single gene, a missing observation for an individual is accommodated for by use of a penetrance function that equals 1 for each genotype.
- Identification of several groups within founder individuals was allowed for, and the procedure to sample allele frequency in a single founder population was then extended to sample allele frequencies in each founder group separately.

For general use of the package, more extensions were made, allowing, for instance, for categorical nongenetic effects to be random and for repeated measurements. These extensions are not described in detail, not being relevant for the Gibbs sampler implementations used in the analyses presented here.

In the software package, alleles for the single gene

are referred to by labels 1 and 2 and the additive effect at the single locus a is not restricted to be positive, so that these allele labels will not be unique. Unique labels to identify alleles and, for instance, allele frequencies, are the Low and High labels as defined in MATERIALS AND METHODS. Correspondence between the two sets of labels is obtained as follows: for $a \geq 0$, label 1 corre-

sponds to Low and label 2 corresponds to High; for $a < 0$, label 2 corresponds to Low and label 1 corresponds to High. Using the so obtained Low and High labels, unique inferences could subsequently be made on allele frequencies, genotypes and genotype frequencies. For the additive effect at the single locus, only the absolute value was considered.