Original article

Likelihood and Bayesian analyses reveal major genes affecting body composition, carcass, meat quality and the number of false teats in a Chinese European pig line

Marie-Pierre Sanchez[a]*, Jean-Pierre Bidanel[a],
Siqing Zhang[a], Jean Naveau[b],
Thierry Burlot[b], Pascale Le Roy[a]

[a] Institut national de la recherche agronomique,
Station de génétique quantitative et appliquée, 78352 Jouy-en-Josas Cedex, France
[b] PEN AR LAN, BP 3, 35380 Maxent, France

Abstract – Segregation analyses were performed using both maximum likelihood – *via* a Quasi Newton algorithm – (ML-QN) and Bayesian – *via* Gibbs sampling – (Bayesian-GS) approaches in the Chinese European *Tiameslan* pig line. Major genes were searched for average ultrasonic backfat thickness (ABT), carcass fat (X2 and X4) and lean (X5) depths, days from 20 to 100 kg (D20100), Napole technological yield (NTY), number of false (FTN) and good (GTN) teats, as well as total teat number (TTN). The discrete nature of FTN was additionally considered using a threshold model under ML methodology. The results obtained with both methods consistently suggested the presence of major genes affecting ABT, X2, NTY, GTN and FTN. Major genes were also suggested for X4 and X5 using ML-QN, but not the Bayesian-GS, approach. The major gene affecting FTN was confirmed using the threshold model. Genetic correlations as well as gene effect and genotype frequency estimates suggested the presence of four different major genes. The first gene would affect fatness traits (ABT, X2 and X4), the second one a leanness trait (X5), the third one NTY and the last one GTN and FTN. Genotype frequencies of breeding animals and their evolution over time were consistent with the selection performed in the *Tiameslan* line.

segregation analysis / likelihood / Bayesian / major gene / pig

---

* Correspondence and reprints
E-mail: sanchez@dga2.jouy.inra.fr

1. INTRODUCTION

Many quantitative trait loci have been identified in pigs with the use of molecular markers [1], leading in a few cases to a causal mutation, as for instance in the case of the *RN* gene [18]. Yet, searching for individual genes using molecular markers is an expensive method, which requires well-planned designs. Segregation analysis, which only uses phenotypic observations, is much less expensive and is complementary to molecular analyses. Indeed, phenotypic analyses only require computing time and can thus be performed on large routinely collected phenotypic data sets, especially from composite lines in which single genes are likely to be segregating.

The composite *Tiameslan* line, which was created by crossing *Laconie* sows and *Meishan* × *Jiaxing* boars, appears to be an interesting population for this purpose. Indeed, genes with major effects on Napole technological yield [14] and backfat thickness [15] have been evidenced in the *Laconie* line. Additionally, particularly high heritability values have been obtained for backfat thickness and the number of total and good teats [25].

A mixed inheritance model, where a major locus effect is added to the classical polygenic variation, is usually constructed to search for major genes. For inference in such a model, maximum likelihood and Bayesian segregation analyses have been successively developed. The maximum likelihood (ML) approach was first used in the human genetics field [4]. Its adaptation to animal genetics has required approximations such as ignoring dependencies between families [13] because animal pedigrees generally contain many loops due to the use of multiple matings. All relationships within a pedigree can now be taken into account using a Monte Carlo Markov chain (MCMC) algorithm [5], such as the Gibbs sampler (GS), generally in a Bayesian inference framework (Bayesian-GS). The GS algorithm was adapted to segregation analysis by Guo and Thompson [7] in order to solve computing problems in complex pedigrees. Later, Janss *et al.* [9] developed a Bayesian-GS approach and a computer software for segregation analyses in livestock species.

Both ML and Bayesian approaches were first developed for normally distributed traits. Elsen and Le Roy however [3] have shown in the case of ML methodology that the use of normality assumptions for discrete traits considerably increase the test statistic values and may therefore lead to the false inference of a major gene. They also showed that the adaptation of ML to discrete variables assuming an underlying normal distribution with a threshold model greatly improves the validity of the test statistics.

The aim of this study was to investigate the existence of major genes affecting false and good teat number and some growth, carcass and meat quality traits in the *Tiameslan* line applying both ML – *via* a Quasi Newton algorithm – (ML-QN) and Bayesian – *via* a GS algorithm – (Bayesian-GS) methods. All traits

were first handled assuming they were normally distributed. The number of false teats was then treated as a discrete trait using a threshold model with ML methodology.

2. MATERIALS AND METHODS

2.1. Animals and measurements

The *Tiameslan* line, developed at the *Pen Ar Lan* nucleus herd of Maxent (Ille-et-Vilaine, France), originated from a cross between sows from the *Laconie* line and Chinese *Meishan × Jiaxing* F1 boars. The breeding company used 55 multiparous sows and 21 boars as founder animals. The data analysed in the present study were composed of 14 generations produced from 1983 to 1996. More details on the *Tiameslan* line can be found in Zhang *et al.* [25]. All animals were weighed at weaning and at the beginning of the test period (at 4 and 8 weeks of age, respectively). At the end of the test period, weight, backfat thickness and the numbers of false and good teats were recorded for all pigs. The teats were classified as false when they were inverted or atrophied. Backfat thickness was measured on each side of the spine at the shoulder, the last rib and the hip joint. Breeding animals were mainly selected on an index combining days from 20 to 100 kg live weight and average backfat thickness. In addition, some selection was performed on teat number (by culling animals carrying false teats) and litter size as described by Zhang *et al.* [25]. The pigs not retained for breeding were slaughtered in a commercial slaughterhouse and measured for Napole technological yield as proposed by Naveau *et al.* [19] until 1990. Carcass fat and lean depths were measured with a "Fat-O-Meater" probe and recorded from 1988 to 1991.

2.2. Traits analysed

Major gene detection was performed for nine different traits: average backfat thickness (ABT = mean of the 6 ultrasonic backfat thickness measurements), carcass fat depth (X2) measured between the 3rd and 4th lumbar vertebrae and carcass fat (X4) and lean (X5) depths measured between the 3rd and 4th last ribs; days from 20 to 100 kg (D20100) defined as the difference between age at 100 kg and at 20 kg, adjusted for weight and age [25]; Napole technological yield (NTY) measured as described by Naveau *et al.* [19]; numbers of good (GTN) and false (FTN) teats, as well as total teat number (TTN = GTN + FTN).

In order to avoid potential bias due to heterosis effects, the performance of founder and F1 animals were discarded. In addition, only sire families with more than 20 offspring were considered in the analyses. The percentage of data removed from the initial data set was 8.5% for X2, X4 and X5, 10.7% for TTN, GTN, FTN, ABT and D20100 and 34% for NTY.

2.3. Data adjustment and transformation

2.3.1. Non-genetic effects

Environmental effects were tested using the General Linear Model procedure of SAS® [22]. A combined sex * batch effect was defined and tested for all traits except NTY where slaughter day was considered as the contemporary group effect. The traits were also adjusted for weight at the start of the test (D20100), at the end of the test (ABT) or for carcass weight (X2, X4 and X5) by including them as linear covariates in the model. All the effects tested were highly significant ($P < 0.001$) for all traits except for X5 where the contemporary group effect only reached a 5% significance level. All the effects investigated were hence kept as adjustment factors. For numerical reasons due to the large number of fixed effect levels (212 and 125 levels for sex * batch and slaughter day, respectively), estimates of the sex * batch and slaughter day effects could not be obtained jointly with the other parameters. The data were thus pre-adjusted for these effects before segregation analyses.

2.3.2. Box-Cox transformation

Additionally, in order to remove skewness that may lead to the false inference of a major gene, the data were transformed using a Box-Cox transformation [17], *i.e.*:

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

where r is a scale parameter to ensure that $(x/r + 1)$ is always positive and p is a power parameter. The power parameter was estimated jointly with the other parameters in ML analyses, whereas the data were transformed before being analysed for genetic parameter estimation and Bayesian analyses. Major gene effects presented later were back-transformed to the original scale using an inverse Box-Cox transformation.

2.4. Estimation of genetic parameters

Genetic parameters of ABT, X2, X4 and X5, were estimated (assuming polygenic inheritance) using restricted maximum likelihood methodology applied to a multivariate animal model with the 4.2.5 version of VCE software [20]. The model included the additive genetic value of each animal and common birth litter as random effects in addition to the fixed effects and covariates described in paragraph 2.3.1. Including D20100 in the analyses was not considered as necessary, since it had previously been shown [25] to have low genetic relationships with carcass composition (or with backfat thickness).

2.5. Major gene detection

2.5.1. Model

The major gene was defined as an autosomal biallelic (*A* and *B*) locus with Mendelian transmission probabilities. In the presence of two alleles *A* and *B*, with probabilities P_A and $P_B = 1 - P_A$, 3 genotypes *AA*, *AB* and *BB* (coded 1, 2 and 3 respectively) can be encountered. A given animal has the genotype *g* (*g* = 1, 2 or 3) with a probability P_g. The vector of phenotypic values **Y** was modelled as:

$$\mathbf{Y} = \mathbf{ZW\mu} + \mathbf{ZU} + \mathbf{E} \tag{1}$$

where **μ** is the vector of genotypic means ($\mu - a$, $\mu + d$, $\mu + a$) associated respectively to the major gene genotypes *AA*, *AB* and *BB*, **U** is the vector of polygenic genetic values and **E** is the vector of residuals; **Z** is an incidence matrix relating genetic effects to observations and **W** is a matrix containing the genotype of each individual. Distributional assumptions for **U** and **E** were $\mathbf{U} \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$, where **A** is the numerator relationship matrix and σ_u^2 is the polygenic variance and $\mathbf{E} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$ where σ_e^2 is the error variance. Polygenic heritability was calculated as $h_{pol}^2 = \sigma_u^2/[\sigma_u^2 + \sigma_e^2]$.

The presence of a major gene was tested under this mixed inheritance model using two different approaches. The first approach was based on the comparison of likelihoods maximised under polygenic and mixed inheritance models [4]. In the second one, statistical inference was based on a Bayesian approach computing marginal posterior densities of the unknown mixed model parameters *via* Gibbs Sampling [9]. In this second approach, computations were performed considering all relationships in the pedigree, whereas ML analyses assumed that data originated from independent families [13]. Under this assumption, only relationships within half- and full-sib families were taken into account in **A**.

2.5.2. Maximum likelihood approach *via a Quasi Newton algorithm (ML-QN)*

The major gene existence was tested comparing the polygenic heredity model (null hypothesis *H*0) to the mixed heredity model (general hypothesis *H*1). The test statistics is the likelihood ratio $l = -2 \ln \dfrac{M0}{M1}$ where *M*1 and *M*0 are the likelihoods under *H*1 and *H*0, respectively.

The sample was assumed to be a set of *n* sire families ($i = 1, \ldots, n$) with m_i mates for sire *i* ($j = 1, \ldots, m_i$) and l_{ij} measured offspring for dam *ij*

$(k = 1, \ldots, l_{ij})$. Following the model (1), $M1$ can then be written:

$$M1 = \prod_{i=1}^{n} \sum_{g_i=1}^{3} p_{g_i} \int_{u_i} f(u_i) f(y_i | u_i, g_i)$$

$$\times \prod_{j=1}^{m_i} \sum_{g_{ij}=1}^{3} p_{g_{ij}} \int_{u_{ij}} f(u_{ij}) f(y_{ij} | u_{ij}, g_{ij})$$

$$\times \prod_{k=1}^{l_{ij}} \sum_{g_{ijk}=1}^{3} P(g_{ijk} | g_i, g_{ij}) f(y_{ijk} | u_i, u_{ij}, g_{ijk}) du_{ij} du_i$$

with:

$$f(u_i) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{1}{2} \frac{u_i^2}{\sigma_u^2}\right), \quad f(u_{ij}) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{1}{2} \frac{u_{ij}^2}{\sigma_u^2}\right),$$

$$f(y_i | u_i, g_i) = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2} \frac{(y_i - u_i - \mu_{g_i})^2}{\sigma_e^2}\right),$$

$$f(y_{ij} | u_{ij}, g_{ij}) = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2} \frac{(y_{ij} - u_{ij} - \mu_{g_{ij}})^2}{\sigma_e^2}\right)$$

and

$$f(y_{ijk} | u_i, u_{ij}, g_{ijk})$$

$$= \frac{1}{\sqrt{2\pi(\sigma_e^2 + \sigma_u^2/2)}} \exp\left(-\frac{1}{2} \frac{\left(y_{ijk} - (u_i + u_{ij})/2 - \mu_{g_{ijk}}\right)^2}{\sigma_e^2 + \sigma_u^2/2}\right)$$

and $M0$ was defined as:

$$M0 = \prod_{i=1}^{n} \int_{u_i} f(u_i) f(y_i | u_i) \prod_{j=1}^{m_i} \int_{u_{ij}} f(u_{ij}) f(y_{ij} | u_{ij})$$

$$\times \prod_{k=1}^{l_{ij}} f(y_{ijk} | u_i, u_{ij}) du_{ij} du_i.$$

FTN was additionally submitted to a segregation analysis with a threshold model assuming that Y is the observed realisation of an underlying normal distribution Z [3]. For a given animal i, the value of y_i is s, if z_i is within the interval $[\lambda_{s-1}; \lambda_s]$ with λ being thresholds, which are estimated jointly with the other parameters. The penetrance function then becomes:

$$f(y_i | u_i, g_i) = \int_{\lambda_{y_i-1}}^{\lambda_{y_i}} \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left[-\frac{1}{2} \frac{(z_i - u_i - \mu_{g_i})^2}{\sigma_e^2}\right] dz_i.$$

Seven parameters were thus estimated ($\mu_1, \mu_2, \mu_3, \sigma_u, \sigma_e, P_{AA}$ and P_{AB}) under $H1$ whereas three parameters were estimated (μ_0, σ_u and σ_e) under $H0$. Maximisation of the likelihoods was made using a quasi-Newton algorithm (E04JYF) of the NAG Fortran library. We supposed that the likelihood ratio l was asymptotically distributed according to a χ^2-distribution with 4 degrees of freedom [13].

2.5.3. The Bayesian approach via a Gibbs sampling algorithm (Bayesian-GS)

The Gibbs sampling algorithm was used for inference in the mixed inheritance model (1) with the MaGGic software package developed by Janss *et al.* [9]. The relationship matrix of the full pedigree was used in the analyses. Marginal posterior densities of a, d, P_A, σ_u^2 and σ_e^2 were estimated and the genotypic variance due to the major gene was computed as: $\sigma_m^2 = 2P_A P_B[a + d(P_B - P_A)]^2 + (2P_A P_B d)^2$ with $P_B = 1 - P_A$. In addition, the proportions of the phenotypic variance due to polygenic effects $[R_u = \sigma_u^2/(\sigma_u^2 + \sigma_m^2 + \sigma_e^2)]$ and to major gene effects $[R_m = \sigma_m^2/(\sigma_u^2 + \sigma_m^2 + \sigma_e^2)]$ were computed. Uniform prior distributions were assumed in the range $(-\infty; +\infty)$ for genotypic values, in the range $[0; +\infty)$ for the variance components and in the range $[0; 1]$ for the allele frequencies. As shown by Hobert and Casella [8], uniform prior distributions lead to proper posterior distributions in the case of linear models. This may not be strictly the case with mixed inheritance models, but we considered that it did not change things much from an operational viewpoint and that the results remained valid.

Gibbs sampler

A trial Gibbs chain of 10 000 iterations was run for each trait and evaluated using the Gibbsit programme [21] to determine the burn-in period (b) and the thinning interval (k). The highest values obtained for b and k (420 and 167, respectively) were increased to 1000 and 500, respectively, and retained as minimum values for all the parameters. In estimation runs, convergence was improved by using the relaxation of allele transmission probabilities to slightly non-Mendelian transmission [23], with only Mendelian samples retained for inference as described by Janss *et al.* [10]. Three chains with different starting values for polygenic and error variances were run per trait. For every chain, 10, 30 or 50% of the phenotypic variance was assigned to the polygenic variance and the remaining part was assigned to error variance. The same starting values were used in the three chains for the other parameters, *i.e.* zero for polygenic and major gene additive and dominance effects and 0.5 for allele frequencies (all the genotypes were initialised as heterozygous). Chain lengths required for convergence were about 25 000 for ABT, X2 and X4; 40 000 for NTY, GTN and FTN and 75 000 for X5.

Post-Gibbs inference

Convergence of the Gibbs sampler was assessed using an analysis-of-variance. For each trait, a chain effect was tested for a, d, P_A, σ_e^2, σ_u^2 and σ_m^2 and convergence was considered as reached when a non-significant chain effect ($> 1\%$) was obtained. Monte Carlo standard errors were computed as described by Sorensen *et al.* [24]. Marginal posterior densities of parameters or functions of parameters were constructed using an average shifted histogram available in the "lash" tool [9]. Means and standard deviations of the posterior distributions were calculated from Gibbs samples.

3. RESULTS

3.1. Trait distributions

The pedigree structure, as well as the means and standard deviations of the nine traits analysed are given in Table I. The size and number of sire families were greater for traits measured on living animals than for carcass traits. All traits appeared as moderately to highly skewed. This was particularly true for GTN and FTN (Fig. 1), whose skewness coefficients reached -2.4 and 5, respectively. Skewness coefficients for the other traits ranged from 0.14 to 1.1. These figures clearly justify the use of the Box-Cox transformation to increase the robustness of the segregation analyses.

Table I. Number of animals, mean and phenotypic standard deviation of the nine traits studied.

Trait [1]	Number of			Mean	Standard deviation
	Sires	Dams	Offspring		
ABT (mm)	114	1255	9231	11.0	2.42
X2 (mm)	22	286	1068	16.2	3.94
X4 (mm)	22	286	1071	16.1	4.00
X5 (mm)	22	288	1100	47.3	7.65
D20100 (d)	114	1255	9231	114	8.99
NTY (%)	35	271	1336	91.3	5.97
GTN	114	1255	9223	15.1	1.93
FTN	114	1255	9223	0.38	1.49
TTN	114	1255	9223	15.4	1.21

[1] ABT = average backfat thickness; X2 and X4 = carcass fat depths; X5 = carcass lean depth; D20100 = days from 20 to 100 kg; NTY = Napole technological yield; GTN = good teat number; FTN = false teat number; TTN = total teat number.
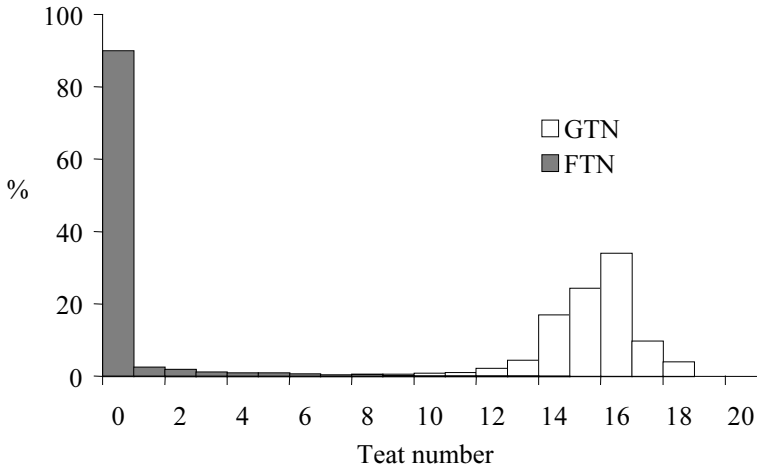
Figure 1. Distribution of good teat number (GTN) and false teat number (FTN).

Table II. Genetic correlation $\pm$ standard error (above the diagonal), heritability $\pm$ standard error (diagonal) and residual correlation $\pm$ standard error (below the diagonal) estimated by VCE for carcass fat depths (X2 and X4), carcass lean depth (X5), and average backfat thickness (ABT).

	X2	X4	X5	ABT
X2	0.46 ± 0.05	0.93 ± 0.02	-0.29 ± 0.07	0.91 ± 0.03
X4	0.75 ± 0.02	0.61 ± 0.05	-0.27 ± 0.05	0.97 ± 0.01
X5	-0.05 ± 0.07	0.03 ± 0.05	0.31 ± 0.04	-0.45 ± 0.05
ABT	0.71 ± 0.03	0.72 ± 0.01	0.17 ± 0.05	0.61 ± 0.06

3.2. Genetic parameters of fatness and lean traits

Genetic parameter estimates for fatness and leanness traits revealed strong genetic correlations between ABT, X2 and X4 (from 0.91 to 0.97), whereas genetic relationships between X5 and fatness traits were much lower, from -0.45 to -0.27 (Tab. II).

3.3. ML-QN approach

3.3.1. Continuous trait analyses

All traits were first analysed assuming that they were normally distributed after Box-Cox transformation. The mixed inheritance model had a much higher likelihood than the purely polygenic model for all traits except TTN and D20100. For these latter traits, the likelihood ratio values were 0 and 3,

Table III. ML-QN results: parameter estimates under a mixed transmission model (*H*1), likelihood ratio value (*l*) and corresponding probability (*P*).

Parameter [1]	a	d	σ_u^2	h_{pol}^2	P_{AA}	P_{AB}	l	$P(l < \chi_4^2)$
Trait [2]								
ABT (mm)	1.7	−2.9	1.0	0.41	0.35	0.65	392	$< 10^{-11}$
X2 (mm)	2.6	−2.6	2.9	0.42	0.29	0.67	31	4×10^{-6}
X4 (mm)	2.2	−2.2	2.3	0.26	0.43	0.38	14	9×10^{-3}
X5 (mm)	5.3	−5.3	19	0.52	0.35	0.61	11	2×10^{-2}
NTY (%)	5.6	−5.6	3.1	0.15	0.60	0.40	50	3×10^{-10}
GTN	2.8	2.9	0.66	0.36	0.00	0.41	737	$< 10^{-11}$
FTN	2.4	−2.4	0.13	0.37	0.59	0.41	4145	$< 10^{-11}$

[1] Parameters under *H*1: a = additive effect; d = dominance effect; σ_u^2 = polygenic variance; h_{pol}^2 = polygenic heritability; P_{AA} and P_{AB} = genotype frequencies of *AA* and *AB*, respectively; [2] see legend of Table I.

respectively, *i.e.* far below the 5% threshold ($\chi_{0.05;4}^2 = 9.5$). The other traits were found to be influenced by a major gene with partial (ABT and GTN) or complete dominance. Yet, it should be noted that likelihood ratio values considerably varied according to the trait, from 11 for X5 to 4145 for FTN (Tab. III).

Major gene effects were rather similar for carcass fatness traits (ABT, X2 and X4), with a dominant allele associated with low values, *i.e.* improved body composition. The mean difference between homozygotes was estimated to be 3.4, 5.1 and 4.4 mm (*i.e.*, 1.6, 2.0 and 1.5 phenotypic standard deviations) respectively, for ABT, X2 and X4. The dominant allele also had favourable effects for GTN and FTN. The animals with a copy of the dominant allele had an average of about 5 more (less) good (false) teats than recessive homozygous animals. These effects represented 4.1 and 7.7 phenotypic standard deviations of GTN and FTN, respectively. Conversely, the major genes evidenced for X5 and NTY had unfavourable dominant alleles. The difference between altern-ative homozygotes for X5 was 1.8 phenotypic standard deviations (10.6 mm). The animals carrying a copy of the dominant allele for NTY had, on aver-age, an 11.2% lower NTY value (*i.e.*, a decrease of 2.5 phenotypic standard deviations).

Estimated frequencies of the favourable genotype in breeding animals were 100, 96 and 81%, respectively, for ABT, X2 and X4. For X5, only 4% of the breeding animals had a favourable genotype. All breeding animals had at least one copy of the dominant alleles decreasing NTY and FTN and increasing GTN.

3.3.2. Discrete trait analyses

The phenotypic distribution of GTN was almost normal. In contrast, the FTN distribution showed a strong deviation from normality (Fig. 1): FTN only took a limited number of values, with a very high frequency (90%) of the zero value. Additional analyses were thus carried out treating FTN as a discrete trait using a threshold model. Adjusted trait values were divided into five classes (≤ -2, -1, 0, 1 and ≥ 2) leading to the estimation of four threshold values. A high likelihood ratio value ($l = 726$; $P(l < \chi_4^2) < 10^{-10}$) was obtained and suggests the presence of a major gene for FTN. The difference between homozygous genotypes was estimated to be 3.8 false teats.

3.4. Bayesian-GS approach

3.4.1. Gibbs chain convergence

The average characteristics of Gibbs chains for each trait are given in Table IV. The percentage of Mendelian samples ranged from 0.65 to 3.8% depending on the trait. Values of the relaxation parameter to obtain the mentioned rates of non-Mendelian samples ranged from 3×10^{-5} to 3×10^{-3}. Among all Mendelian samples obtained, every 10 to 36th was retained for inference while the lowest burn-in period and thinning interval retained were equal to 1111 and 691, respectively. Chain length ranged from 69 140 to 175 560 iterations, *i.e.* were longer than the lowest value

Table IV. Bayesian-GS results: Gibbs chain characteristics (means of the three chains per trait).

	Burn in	Spacing between iterations retained	Chain length	P_{nmt} [2] initial-final	% of Mendelian samples	Spacing between Mendelian samples	MCse [3]
Trait [1]							
ABT	1134	1770	175 560	$4 \times 10^{-2} - 4 \times 10^{-4}$	1.0	17	0.007
X2	2552	1494	161 777	$5 \times 10^{-2} - 1 \times 10^{-3}$	1.9	30	0.009
X4	1850	691	69 140	$5 \times 10^{-1} - 3 \times 10^{-3}$	2.9	20	0.04
X5	1657	875	95 404	$5 \times 10^{-1} - 3 \times 10^{-3}$	3.8	36	0.07
NTY	3138	1137	126 989	$5 \times 10^{-2} - 3 \times 10^{-3}$	1.1	14	0.02
GTN	1111	1565	155 091	$4 \times 10^{-2} - 3 \times 10^{-5}$	0.65	10	0.005
FTN	1319	1020	101 314	$3 \times 10^{-3} - 4 \times 10^{-4}$	2.0	18	0.002

[1] See legend of Table I; [2] probability of non Mendelian allele transmission; [3] Monte Carlo standard error.

required for convergence (Tab. IV). One hundred Mendelian samples were retained per chain, so that a total of 300 samples were available per trait. Monte Carlo standard errors were low for all trait × parameter combinations. Independence of the samples was tested by an analysis-of-variance. A significant chain effect ($P = 8 \times 10^{-4}$) was observed for a single trait × parameter combination, *i.e.* the dominance effect for ABT. Convergence was considered to be reached for all the other parameters and the samples from the three chains were combined to estimate features of marginal posterior distributions.

3.4.2. Inference in a mixed inheritance model

Estimated posterior marginal distributions of polygenic and major gene variance ratios are shown in Figure 2. The highest posterior density regions at 95% ($HPD_{95\%}$) of model parameters (Tab. V) strongly suggest the presence of major gene effects for ABT, X2, NTY, GTN and FTN. Less clear results were obtained for X4 and X5 since the $HPD_{95\%}$ of R_m were far from zero for all traits except X4 and X5. Table V also shows marginal posterior means and standard deviations of mixed inheritance model parameters. Partial or full dominance effects were obtained for all major genes detected. The dominant allele was associated with decreased ABT, X2, NTY and FTN values and increased GTN values. It was hence favourable for all traits except NTY. For ABT and X2, the differences between homozygous animals were 2.6 and 10.9 mm, *i.e.* 1.4 and 3.8 phenotypic standard deviations, respectively. Recessive homozygotes had, on average, 10.6% (2.3 phenotypic standard deviations) lower NTY values than dominant homozygous animals. Concerning GTN and FTN, the difference between homozygotes was estimated to 2.6 and 2.4 teats, respectively (1.8 and 3.6 phenotypic standard deviations, respectively).

Genotype frequencies were estimated from marginal posterior means of allele frequencies assuming Hardy-Weinberg equilibrium. Most of the breeding animals had a favourable genotype, except for NTY, where 97% of the animals had an unfavourable genotype.

The evolution of allele frequencies over time was also estimated from individual genotype probabilities. This estimation was carried out in the population of reproductive females, which provides a reasonable compromise between population size and the precision of individual genotype probabilities. Allele frequencies were only estimated for ABT, GTN and FTN because the number of measurements was too low for the other traits. The frequency of the favourable alleles regularly increased over generations for the three traits. It was intermediate (from 0.43 for ABT to 0.50 for GTN and FTN) in the 2nd generation and close to one after 12 years of selection.
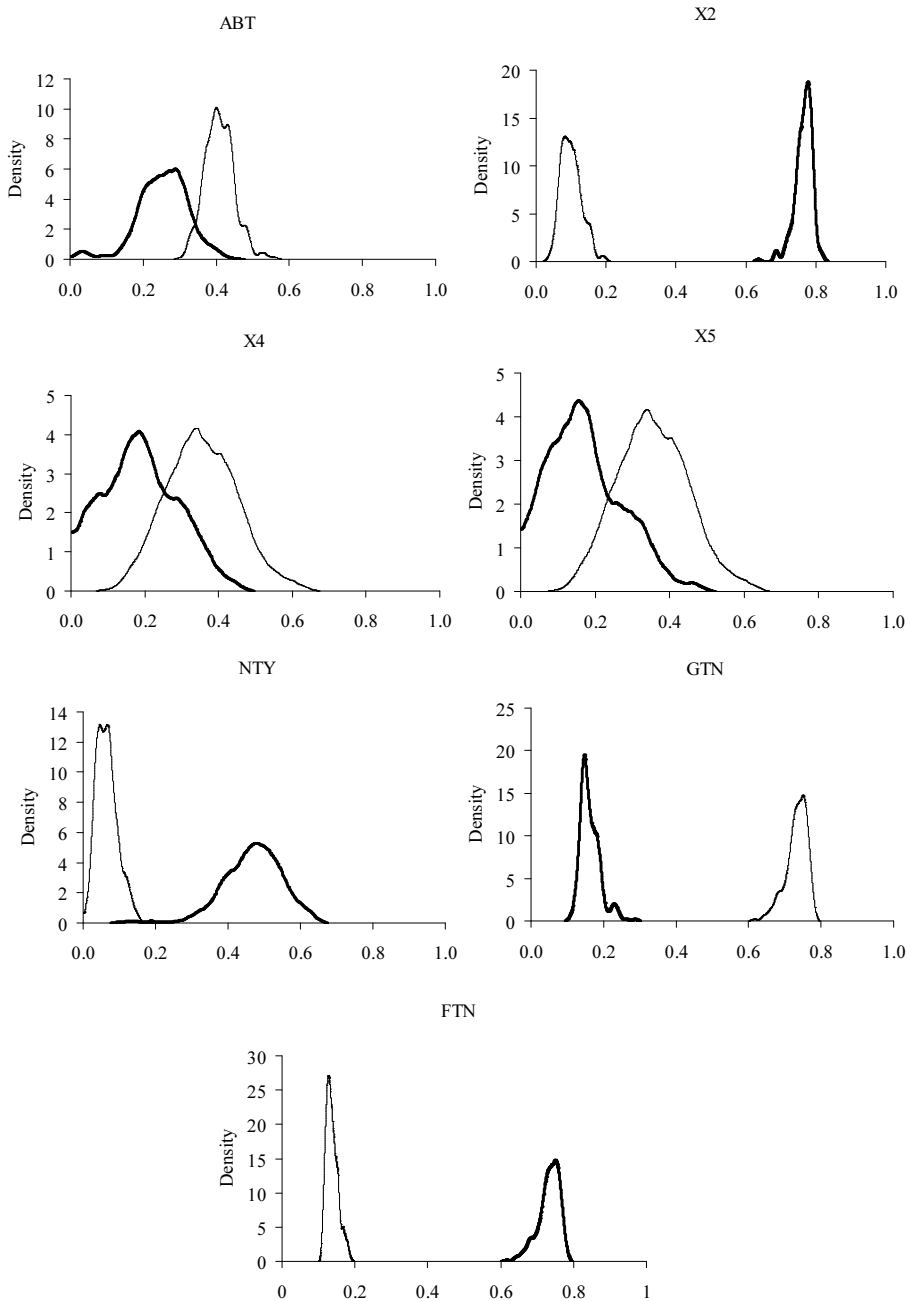
ABT

X2

X4

X5

NTY

GTN

FTN

Figure 2. Marginal posterior distributions of polygenic (——) and major gene (——) variance ratios for average backfat thickness (ABT), carcass fat depths (X2 and X4) and carcass lean depth (X5), Napole technological yield (NTY), good teat number (GTN) and false teat number (FTN).

Table V. Bayesian-GS results: marginal posterior means (mpm), marginal posterior standard deviations (mpsd) and highest posterior density region at 95% (HPD$_{95\%}$) of mixed inheritance model parameters.

Trait [1]	a [1]			d			P_A			R_u			R_m		
	mpm	mpsd	HPD$_{95\%}$	mpm	mpsd	HPD$_{95\%}$	mpm	mpsd	HPD$_{95\%}$	mpm	mpsd	HPD$_{95\%}$	mpm	mpsd	HPD$_{95\%}$
ABT	1.3	0.21	1.0/1.5	−1.7	0.25	−2.1/−1.2	0.41 *	0.072	0.27/0.56	0.43	0.085	0.32/0.49	0.24	0.14	0.13/0.41
X2	5.4	0.33	4.8/6.1	−5.1	0.34	−5.7/−4.4	0.68	0.083	0.52/0.84	0.12	0.057	0.05/0.16	0.73	0.061	0.71/0.81
X4	1.1	0.67	0.003/2.3	0.63	2.5	−3.5/4.4	0.52	0.14	0.22/0.76	0.37	0.18	0.16/0.54	0.25	0.14	0.001/0.37
X5	2.3	1.3	0.006/4.6	0.44	4.3	−6.7/7.1	0.52	0.14	0.26/0.79	0.29	0.17	0.06/0.41	0.26	0.15	0.0005/0.36
NTY	5.3	0.94	3.5/7.3	−4.3	1.0	−6.3/−2.5	0.82	0.064	0.63/0.93	0.11	0.059	0.01/0.13	0.38	0.17	0.32/0.63
GTN	1.3	0.078	1.1/1.5	2.0	0.19	1.6/2.3	0.38	0.076	0.23/0.54	0.20	0.060	0.12/0.23	0.44	0.12	0.39/0.63
FTN	1.2	0.016	1.1/1.2	−1.3	0.021	−1.3/−1.2	0.54	0.062	0.41/0.67	0.15	0.028	0.11/0.17	0.70	0.056	0.67/0.78

[1] a = additive effect; d = dominance effect; P_A = frequency of allele A; R_u = polygenic variance ratio; R_m = major gene variance ratio; [2] see legend of Table I; * significant chain effect ($P = 8 \times 10^{-4}$).

4. DISCUSSION

The two different approaches used in this study are the most commonly used ones to detect major genes. They differ from several standpoints. In particular, the information given by a marginal posterior distribution is much more comprehensive than the information given by an ML analysis. Inference relies on very different concepts that are detailed by Gianola and Fernando [6] and Blasco [2] and will not be further discussed here. They also differ from an operational standpoint, since ML-QN is a rather fast method, but it requires simple pedigree structures, whereas Bayesian-GS can easily handle complex pedigrees, but is computationally demanding. In our study, even if we cannot strictly compare the two approaches because none of the procedures were completely optimised for our data, the ML-QN approach appeared as much faster than the Bayesian-GS one. For instance, FTN analyses required about 20 min with the first approach *vs.* 16 h of CPU time with the second one. The ML-QN approach could therefore be used more easily in routine analyses than the Bayesian-GS approach.

Generally speaking, ML-QN and Bayesian-GS approaches gave rather consistent results in our study. Both methods led to the assumptions that major genes affecting ABT, X2, NTY, GTN, FTN are existing. Heritability and genotype frequency estimates were also similar with both methods. Estimates of major gene effects were also rather consistent, even if the Bayesian-GS approach generally gave lower values of additive and dominance effects than ML-QN. Thus, ignoring dependencies between families did not dramatically change the results. This result certainly has to be related to the very large family size in the data set analysed. Indeed, most of the genetic information contained in the pedigree concerned the segregation of genes within half-sib families. They probably cannot be extended to the case of pedigrees with smaller family size, such as human pedigrees for instance.

It is well known that segregation analysis is not robust to violations to model assumptions, such as the normality of residuals. We tried to minimise the impact of the non-normality of residuals by using a Box-Cox transformation. Yet, departures from other assumptions, such as the homogeneity of the distribution of residuals among families, may also lead to the false inference of a major gene. The results from segregation analyses should therefore be considered as first indications of the presence of the major gene, which have then to be confirmed by analyses using molecular tools. Nevertheless, estimates of genetic correlations and major gene effects tend to suggest that four major genes might be segregating in the Tiameslan population. Two of them would affect carcass composition, the first one with pleiotropic effects on ABT, X2 and X4, and the second one affecting X5. Indeed, the three fatness traits are genetically very closely related, as shown by very high genetic correlations,

and the alleles detected for ABT, X2 and X4 have many similarities. This gene may be the same as that previously described in the *Laconie* line, one of the *Tiameslan* founder lines [15], since genotype effects on ABT were very similar in both populations. Conversely, the low to medium genetic correlations between fatness traits and X5, as well as the major gene effects on this latter trait, tend to show that the gene involved differs from the previous one.

The third major gene affecting NTY has many similarities with the *RN* gene detected in the *Laconie* line by Le Roy *et al.* [14], and it can be assumed that the observed effects are due to the RN^- mutation [18].

Clear evidence for the segregation of a fourth major gene influencing the numbers of false and good teats was found. Even if the discrete nature of the trait makes it more likely to find false positive results, the very high likelihood ratios combined with the analyses taking into account the discrete nature of the trait makes the existence of this gene highly probable. This major gene influencing the number of false or good teats without any effect on total teat number, had, to our knowledge, never been described before. In particular, its effects are not directly related to the favourable effects of the *Meishan* and, above all, the *Jiaxing* breeds, on teat number [12]. It is not possible to know from our data whether this gene primarily acts on the number of false or the number of good teats. Indeed, it may be argued that it either acts as a genetic defect by increasing the number of teats with irregular forms or as a necessary gene for teats to have a normal shape and become active. It should also be emphasised that some of the teats classified as "false" at the end of the performance test become active after farrowing, as recently shown by Labroue *et al.* [11].

Genotype frequencies in breeding animals and the evolution of the frequency of favourable alleles over time were consistent with the selection performed in the *Tiameslan* line against backfat thickness and the number of false teats. The proportion of breeding animals with a favourable genotype was lower for X2, but it remained close to unity and was consistent with genetic correlation estimates of fatness traits. Even if no selection was carried out for the Napole technological yield before 1990, the very high RN^- allele frequency (about 0.8 with both methods) obtained among breeding animals is somewhat surprising, since it was estimated at only 0.6 in the *Laconie* line in 1986 [14]. One possible explanation is a correlated response to selection against backfat thickness, since the RN^- mutation has been shown to have pleiotropic effects on NTY and ABT [16].

5. CONCLUSION

Based on the results of this study, it can be concluded that the ML approach can be considered as a reliable approach to detect major genes, at least in pedigrees with large sire families. However, Bayesian-GS makes more

comprehensive analyses possible, such as allowing genetic trends (polygenic and at the major locus) over generations to be estimated. Since the ML-QN approach is much less time consuming than the Bayesian-GS, we suggest applying the ML-QN approach to detect major genes and then performing a Bayesian-GS method in order to better characterise the major gene effects detected.

This study allowed to confirm the existence of two major genes affecting, respectively, fatness traits and the Napole technological yield, previously described in the *Laconie* line. In addition, a new major gene affecting the number of false teats was evidenced and a major gene affecting muscle depth was suggested. The next step will then be to localise these genes using an experimental design, *i.e.* a set of informative families for the genes and for available panels of molecular markers. Molecular markers could make it possible to easily identify heterozygous animals carrying the unfavourable recessive allele and would be very helpful to more efficiently eradicate these alleles.

REFERENCES

[1] Bidanel J.P., Rothschild M.F., Current status of quantitative trait locus mapping in pigs, Pig News Inf. 23 (2002) 39N–53N.

[2] Blasco A., The Bayesian controversy in animal breeding, J. Anim. Sci. 79 (2001) 2023–2046.

[3] Elsen J.M., Le Roy P., Detection of major genes and determination of genotypes application to discrete variables, in: Proceedings of the 4th World Congress on Genetics Applied to Livestock Production, Edinburgh, 23–27 July 1990, Vol. 15, pp. 37–49.

[4] Elston R.C., Stewart J., A general model for the genetic analysis of pedigree data, Human Hered. 21 (1971) 523–542.

[5] Gelfand A.E., Smith A.F.M., Sampling based approaches to calculating marginal densities, J. Am. Stat. Assoc. 85 (1990) 398–409.

[6] Gianola D., Fernando R.L., Bayesian methods in animal breeding theory, J. Anim. Sci. 63 (1986) 217–244.

[7] Guo S.W., Thompson E.A., Monte Carlo estimation of mixed models for large complex pedigrees, Biometrics 50 (1994) 417–432.

[8] Hobert J.P., Casella G., Gibbs sampling with improper prior distributions, Technical report BU-1221-M, Biometrics Unit, Cornell University, 1994.

[9] Janss L.L.G., Thompson R., van Arendonk J.A.M., Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations, Theor. Appl. Genet. 91 (1995) 1137–1147.

[10] Janss L.L.G., van Arendonk J.A.M., Brascamp E.W., Bayesian statistical analyses for presence of single genes affecting meat quality traits in a crossed pig population, Genetics 145 (1997) 395–408.

[11] Labroue F., Caugant A., Ligonesche B., Gaudré D., Étude de l'évolution des tétines d'apparence douteuse chez la cochette au cours de sa carrière, J. Rech. Porc. Fr. 33 (2001) 145–150.

[12] Legault C., Caritez J.C., L'expérimentation sur le porc chinois en France. I. Performances de reproduction en race pure et en croisement, Génét. Sél. Évol. 15 (1983) 225–240.

[13] Le Roy P., Elsen J.M., Knott S., Comparison of four statistical methods for detection of a major gene in a progeny test design, Genet. Sel. Evol. 21 (1989) 341–357.

[14] Le Roy P., Naveau J., Elsen J.M., Sellier P., Evidence for a new major gene influencing meat quality in pigs, Genet. Res. 55 (1990) 33–40.

[15] Le Roy P., Elsen J.M., Naveau J., Étude de la variabilité génétique de l'adiposité dans la lignée Laconie, J. Rech. Porc. Fr. 22 (1990) 11–16.

[16] Le Roy P., Elsen J.M., Caritez J.C., Talmant A., Juin H., Sellier P., Monin G., Comparison between the three porcine *RN* genotypes for growth, carcass composition and meat quality traits, Genet. Sel. Evol. 32 (2000) 165–186.

[17] MacLean C.J., Morton N.E., Elston R.C., Yee S., Skewness in commingled distributions, Biometrics 32 (1976) 695–699.

[18] Milan D., Jeon J.T., Looft C., Amarger V., Robic A., Thelander M., Rogel-Gaillard C., Paul S., Iannuccelli N., Rask L., Ronne H., Lundström K., Reinsch N., Gellin J., Kalm E., Le Roy P., Chardon P., Andersson L., A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle, Science 288 (2000) 1248–1251.

[19] Naveau J., Pommeret P., Lechaux P., Proposition d'une méthode de mesure du rendement technologique "la méthode Napole", Techni-porc 8 (1985) 7–13.

[20] Neumaier A., Groeneveld E., Restricted maximum likelihood estimation of covariances in sparse linear models, Genet. Sel. Evol. 30 (1998) 3–26.

[21] Raftery A.E., Lewis S.M., Implementing MCMC, in: Gilks W.R., Richardson S., Spiegelhalter D.J. (Eds.), Markov chain Monte Carlo in practice, Chapman & Hall, London, 1996, pp. 115–130.

[22] SAS® Institute, SAS®/STAT, User's Guide: Statistics (Version 6). SAS® Institute Inc., Cary, NC, 1992.

[23] Sheehan N., Thomas A., On the irreducibility of a Markov chain defined on a space of genotype configurations by a sampling scheme, Biometrics 49 (1993) 163–175.

[24] Sorensen D.A., Andersen S., Gianola D., Korsgaard I., Bayesian inference in threshold models using Gibbs sampling, Genet. Sel. Evol. 27 (1995) 229–249.

[25] Zhang S., Bidanel J.P., Burlot T., Legault C., Naveau J., Genetic parameters and genetic trends in the Chinese × European *Tiameslan* composite pig line. I. Genetic parameters, Genet. Sel. Evol. 32 (2000) 41–56.