Journal of
Animal Breeding and Genetics

J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

The distribution of QTL additive and dominance effects in porcine F2 crosses

J. Bennewitz[1] & T.H.E. Meuwissen[2]

1 Institute of Animal Husbandry and Breeding, University of Hohenheim, Stuttgart, Germany
2 Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås, Norway

Keywords
Exponential distribution; mixtures of normals; QTL effects.

Correspondence
J. Bennewitz, Garbenstrasse 17, Institute of Animal Husbandry and Breeding, University of Hohenheim, D-70599 Stuttgart, Germany. Tel: +49 711 459 23570; Fax: +49 711 459 23101; E-mail: j.bennewitz@uni-hohenheim.de

Received: 28 April 2009;
accepted: 11 November 2009

Summary

The present study used published quantitative trait loci (QTL) mapping data from three F2 crosses in pigs for 34 meat quality and carcass traits to derive the distribution of additive QTL effects as well as dominance coefficients. Dominance coefficients were calculated as the observed QTL dominance deviation divided by the absolute value of the observed QTL additive effect. The error variance of this ratio was approximated using the delta method. Mixtures of normal distributions (mixtures of normals) were fitted to the dominance coefficient using a modified EM-algorithm that considered the heterogeneous error variances of the data points. The results suggested clearly to fit one component which means that the dominance coefficients are normally distributed with an estimated mean (standard deviation) of 0.193 (0.312). For the additive effects mixtures of normals and a truncated exponential distribution were fitted. Two components were fitted by the mixtures of normals. The mixtures of normals did not predict enough QTL with small effects compared to the exponential distribution and to literature reports. The estimated rate parameter of the exponential distribution was 5.81 resulting in a mean effect of 0.172.

Introduction

Although many studies with the aim of mapping quantitative trait loci (QTL) in livestock species were conducted during the last decades (Khatkar *et al.* 2004), little is known about the distribution of the QTL effects. This is especially true for dominance QTL effects. Having knowledge of the distribution would contribute to the understanding of the genetics of quantitative traits and is therefore of interest in several fields. For example, it would help to assess to which amount the assumptions of the infinitesimal model, which assumes an infinite large number of QTL with an infinite small effect, is violated (Goddard 2001). Particular use of the distribution of additive QTL effects has been made to obtain bias-reduced QTL estimates (Hayes & Goddard 2001; Weller *et al.* 2005), to estimate the number of QTL underlying quantitative trait variation (Otto & Jones 2000; Hayes & Goddard 2001), and to develop priors for QTL mapping and genomic breeding value estimation (Meuwissen *et al.* 2001; Goddard 2009).

Using QTL mapping results from livestock species, Hayes & Goddard (2001) and Weller *et al.* (2005) fitted gamma distributions in order to model QTL effects. The authors found the distributions mostly to be leptocurtic, with many QTL of small and a few QTL of large effect. The gamma distribution was chosen by the authors because of its flexibility with regard to the shape of the distribution. Keightley (1994) argued that the distribution of mutant effects is probably not a gamma but likely a complex

  doi:10.1111/j.1439-0388.2009.00847.x

mixture of distributions, because mutation events fall into several classes with different effects. Chamberlain (2004) argued that it is desirable to use a method that is free of any assumptions with regard to the shape of the distribution. Following this, mixtures of normal distributions (mixtures of normals) (e.g. McLachlan & Peel 2000), are an interesting class of distribution for the modelling of QTL effects. They can be seen as somewhere in between parametric and non-parametric approaches. They are parametric, because they fit normal distributions. The non-parametric characteristic is due to the potentially large number of fitted components, theoretically up to the number of observations (McLachlan & Peel 2000).

The present study used published QTL mapping data from three F2 crosses in pigs for 34 meat quality and carcass traits to derive the distribution of additive QTL effects as well as dominance coefficients, taking the heterogeneous error variances of the observations and the fact that only significant QTL are reported into account. Mixtures of normals as well as gamma and exponential distributions were used.

Material and methods

Data and data editing

Geldermann *et al.* (2003) published additive (y_{add}) and dominance (y_{dom}) QTL effect estimates together with their standard errors. Three F_2 pig crosses derived from European Wild Boar, Meishan and Pietrain crosses were used to conduct a QTL genome scan. The number of animals in the parental lines were 10 (68, 61) for Wild Boar (Meishan, Pietrain). The number of F2 animals were 315 (316, 347) for the Wild Boar × Pietrain (Meishan × Pietrain, Wild Boar × Meishan) cross. In total 46 meat quality and carcass traits were analysed. All animals were housed on one farm and trait measurement took place under standardized conditions (Müller *et al.* 2000). The animals were genotyped for the same set of markers. The three crosses were analysed separately using multi-marker interval mapping implemented in the least-squares methods of Haley *et al.* (1994), see Geldermann *et al.* (2003) for details. Briefly, assuming that the parental lines were fixed for alternative genotypes at putative QTL, for each F_2 individual and each chromosomal position (i.e. each cM) the three QTL genotype probabilities were estimated using the marker data. From these probabilities two regression variables were derived, one for the additive QTL effect and one for the dominant QTL effect. The regression variable for the additive effect was the difference of the two probabilities of being homozygous at the QTL, i.e. pr(QQ) − pr(qq). For the dominant QTL effect it was the probability of being QTL-heterozygous. The phenotypes were adjusted for some fixed effects and were regressed univariately on these two variables, the test statistic was an *F*-value. The analysis was repeated every cM. The null (alternative) hypothesis was that there is no (one) QTL on a particular chromosome for the particular trait being analysed. Chromosomewise critical values were obtained by a permutation test. The position with the highest test statistic on a chromosome was taken as the most likely QTL position. A QTL was reported to be significant if the test statistic at the most likely position exceeded the 5% chromosomewise critical value. The results from these QTL analysis were published separately for each chromosome (see *J. Anim. Breed. Genet.*, 2003, Vol. **120**, Supplemental 1). The y_{add} effect was half the estimated differences between the two homozygous genotypic values. The y_{dom} effect was the estimated deviation of the heterozygous from the midpoint between the two homozygous. In the present study, we scaled all QTL effects by the phenotypic standard deviations of the traits. Müller *et al.* (2000) estimated the LS-means of the trait values in the same F_2 crosses. From the standard errors of these LS-means and the number of F_2 individuals the phenotypic standard deviation of 34 traits were derived (see Appendix A for trait definition). Only the QTL results of these traits were used in this study, resulting in n = 271 QTL across all three crosses (80 from European Wild Boar × Meishan cross, 94 from European Wild Boar × Pietrain cross and 97 from Meishan × Pietrain cross). The histogram of y_{add} and y_{dom} is shown in Figures 1 and 2, respectively.

To test whether the dominance effects, y_{dom}, were real or random deviations, their significance was tested as follows. The null hypothesis was that the effect estimates were multivariate normal distributed with mean zero and variance equal to the error variance (i.e. the squared standard error). The alternative hypothesis was that the absolute mean was greater than zero. The squared effect estimate was divided by its error variance and this ratio was summed up over all n QTL. If the null hypothesis was true, this sum follows a chi-square distribution with n degrees of freedom. The results showed that the mean of the y_{dom} effects was significantly different from zero ($p < 0.001$).

In order to be comparable with other studies (e.g. Caballero and Keightley 1994; Dilda & Mackay

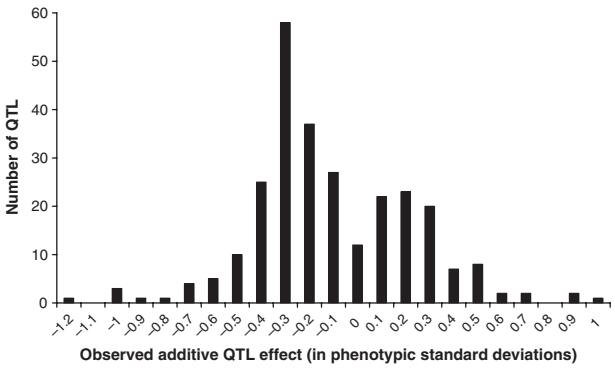

Figure 1 Histogram of observed additive quantitative trait loci (QTL) effects, y_{add} (n = 271).

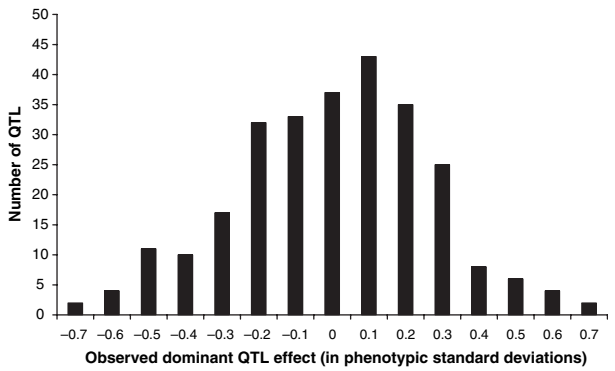

Figure 2 Histogram of observed dominant quantitative trait loci (QTL) effects, y_{dom} (n = 271).

2002) we used in the following the observed dominance coefficient: $y_h = y_{dom}/|y_{add}|$. The error variance of this ratio (SE^2) was approximated by the delta method (e.g. Lynch and Walsh, 1998, page 818) using the error variance of the two estimates and assuming no covariance between them. The latter one is because the covariance was not known. The histogram of y_h is shown in Figure 3.

Fit of mixtures of normals to QTL additive effects and dominance coefficients

We used the following model (the indices 'add' and 'h' are dropped for simplicity)

$$y_i = a_i + e_i, \text{ and}$$

$$\text{var}(y_i) = \text{var}(a_i) + \text{var}(e_i) \quad (1)$$

with y_i the ith observed QTL effect, a_i the ith true QTL effect and e_i the ith error ($i = 1,\ldots n$). Var(a_i) is the variance of the QTL effects we wanted to infer

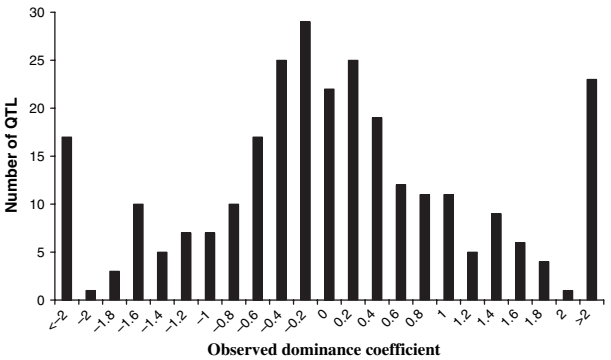

Figure 3 Histogram of the observed dominance coefficients, $y_h = y_{dom}/|y_{add}|$ (n = 271).

(subsequently denoted by σ^2) and var(e_i) is the error variance, which is the square of the standard error of the observed QTL effect (SE_i^2)and this was known from Geldermann *et al.* (2003). Further, it was assumed that each a_i is a realisation from one of g components with mean μ_j ($j = 1,\ldots, g$) and variance σ_j^2. Assuming that it were known from which component a_i was drawn, one could write the complete-data log likelihood as follows

$$\log L_c(\theta) = \sum_{j=1}^{g}\sum_{i=1}^{n} z_{ij} \log\left\{\pi_j \phi(y_i; \mu_j, \sigma_j^2 + \text{SE}_i^2)\right\} \quad (2)$$

where z_{ij} is a known indicator variable indicating to which component y_i belongs. π_j are the mixing proportions, they are non-negative and sum up to one. θ is a vector containing the unknown parameters π_j, μ_j and σ_j^2. This formulation was adapted from McLachlan & Peel (2000), except that the variance in the normal p.d.f. ($\phi(\ldots)$) is the sum of two variance components. However, z_{ij} was unknown and had to be estimated. This led to the following EM algorithm (see e.g. McLachlan & Peel 2000). On the $k + 1$ iteration in the E-step the conditional expectation of z_{ij} given y_i and the current $\theta^{(k)}$ was estimated as

$$E(z_{ij}|y_i, \theta^{(k)}) = \tau_{ij} = \frac{\pi_j \phi(y_i; \mu_j, \sigma_j^2 + \text{SE}_i^2)}{\sum_{h=1}^{g} \pi_h \phi(y_i; \mu_h, \sigma_h^2 + \text{SE}_i^2)} \quad (3)$$

The M-step consisted of the following computations (4 and 5) for estimating the mixing proportions, means and variances of the components

$$\pi_j^{(k+1)} = \sum_i^n \tau_{ij}/n$$

and

$$\mu_j^{(k+1)} = \sum_{i=1}^n \tau_{ij} y_i / \sum_{i=1}^n \tau_{ij}. \quad (4)$$

Because the variance of the g component was the sum of two variance components, the estimation of σ_j^2 differed from the usual EM-algorithm and was done in the M-step using mixed model theory by iterating on the following equation (l indicates the iteration) (see Appendix B for a full derivation)

$$\left(\sigma_j^2\right)^{(k+1),(l+1)} = \frac{\sum_{i=1}^n \tau_{ij}\left(h_j^2\right)^{(k+1),l}\left[\left(h_j^2\right)^{(k+1),l} y_i^2 + \mathrm{SE}_i^2\right]}{\sum_{i=1}^n \tau_{ij}} \quad (5)$$

with $(h_j^2)^{(k+1),l} = [(\sigma_j^2)^{(k+1),l}/((\sigma_j^2)^{(k+1),l} + \mathrm{SE}_i^2)]$. Iteration was until $|(\sigma_j^2)^{(k+1),(l+1)} - (\sigma_j^2)^{(k+1),l}|$ became $<10^{-8}$. Starting value for $(\sigma_j^2)^{(k+1),l}$ was

$$\left(\sigma_j^2\right)^{(k+1),l=1} = \left(\sigma_j^2\right)^* - \sum_{i=1}^n \tau_{ij}\mathrm{SE}_i^2 / \sum_{i=1}^n \tau_{ij},$$

where $(\sigma_j^2)^*$ was the variance of component j estimated with the 'usual' M-step, i.e. ignoring the error variances of the data points. For executing the EM-algorithm the EMMIX software (Peel & McLachlan 1999) was modified according to Equations (2) to (5).

The number of components was determined by the statistical test implemented in EMMIX and is described in McLachlan (1987). The confidence intervals of the parameters were estimated by non-parametric bootstrapping. A total of 200 non-parametric bootstrap samples were generated by sampling with replacement n observations out of the pool of n original observations. The bootstrap samples were analysed with methods shown above. Subsequently, the bootstrap samples were ranked according to the estimates of a parameter and the estimates of the 5th and 195th bootstrap sample were taken as the lower and upper bounds of the support limit. This corresponded to a 95% confidence interval.

The EM-algorithm was applied separately to the additive and dominance data sets. Only significant QTL were used, resulting in a reduced density of y_{add} around zero (Figure 1). We argue that this is due to lower experimental power mapping these QTL and if all QTL were mapped the highest density would be observed at point zero. Hence, it is assumed that the highest density of the true QTL effects is around zero. In order to model this, for each y_{add} a second effect was generated with exactly the same size and same error variance but with an opposite sign. For example, assume an observed QTL effect estimate of 0.3 with an SE of 0.01. The additional generated record belonging to this data point would be (−0.3), also with an SE of 0.01. This kind of 'doubling' could be done because the absolute values of the alternative homozygous genotypes at the QTL are the same by definition (Falconer and Mackay 1996) and the sign of y_{add} is arbitrary, since it depends on which of the QTL alleles is called 'Q' and which is called 'q'. The 'doubling' guaranteed that the mean of all data points in this 'doubled' data set was equal to zero. Subsequently, during the analysis of this data set the mean of each component in the M-step was set equal to zero rather than estimated by Equation (4). By doing this it is guaranteed that the highest density is at zero for each component.

It seemed that the y_h are not affected by the use of only significant QTL, because the highest density of y_h is around zero (Figure 3). Therefore, no additional data editing was done and the means of the components were estimated rather than setting them equal to zero.

Fit of gamma and exponential distribution to QTL additive effects

Only significant published QTL effects were also used by Hayes & Goddard (2001) and they fitted therefore a truncated gamma distribution. We followed this idea and fitted a truncated gamma distribution to the additive effects. Because the gamma distribution is defined only for non-negative values, the absolute values of the additive effects were taken. The gamma distribution with scaling parameter α and shape parameter β of the additive QTL effects was $g(a) = \alpha^\beta a^{\beta-1} e^{-\alpha a}/\Gamma(\beta)$. The mean and the variance is $E(a) = \beta/\alpha$ and $\mathrm{var}(a) = \beta/\alpha^2$, respectively. Following Hayes & Goddard (2001), the density of y_i given the true underlying effect a_i, $n(y_i|a_i)$ is

$$n(y_i|a_i) = \frac{1}{\sqrt{2*\pi}*\mathrm{SE}_i} e^{-\left(\frac{(y_i-a_i)^2}{2*\mathrm{SE}_i^2}\right)} \quad (6)$$

Unlike Hayes & Goddard (2001) we used the SE^2 of each single y_i instead of averaging over the experiment. Although y_i is always positive for the additive QTL effects it is possible that y_i and a_i have opposite

signs. The density function for y_i can therefore be written as (Hayes & Goddard 2001)

$$f(y_i) = \int_0^\infty n(y_i|a_i)g(a)da + \int_o^\infty n(-y_i|a_i)g(a)da \qquad (7)$$

Taking into account the truncation point (c), the probability of observing y_i given α and β is (Hayes & Goddard 2001)

$$P(y_i|\alpha,\beta) = \frac{f(y_i)}{\int_c^\infty f(y)dy}. \qquad (8)$$

The log likelihood for all y_i is $\sum_{i=1}^{n} \ln P(y_i|\alpha,\beta)$. Monte Carlo integration was used to integrate the distributions (Tanner 1993). A grid search was used to find the maximum likelihood estimates of α and β given the data.

Hayes & Goddard (2001) defined the truncation point as the smallest of the significant observed QTL effect. In our study this value would be close to zero (Figure 1) (the QTL with y_{add} close to zero showed a large y_{dom}, therefore they were still significant). Therefore, we determined the truncation point as follows. The shape parameter indicates whether the distribution is L-shaped ($\beta < 1$) or skewed ($\beta > 1$). A special case is $\beta = 1$, resulting in an exponential distribution. We assumed that the highest density is around zero and therefore a skewed distribution is not appropriate. Our truncation point was the lowest y_{add} that resulted in a $\beta = 1$ applying Equations (6) to (8), if all QTL with a smaller y_{add} were deleted from the data set. This reduced data set was analysed applying Equations (6) to (8) but fixing $\beta = 1$ instead of estimating it. Thus, the final distribution we fitted to the additive QTL effects was an exponential distribution with rate parameter $1/\alpha$. The mean and variance of this distribution is $E(a) = 1/\alpha$ and $\text{var}(a) = 1/\alpha^2$, respectively. The confidence interval of α was calculated by bootstrapping as described above.

Results

A y_h >1 (1, 0, −1, <−1) indicates overdominance (complete dominance, additivity, complete recessivity, underdominance). Following this, all kinds of dominance are present in the data (Figure 3), with a substantial proportion of QTL with over- or underdominance. Compared to dominance coefficients from Drosophila experiments (Dilda & Mackay 2002), a much larger proportion show extreme values. However, most of the extreme data points show either a large standard error of y_{add} or of y_{dom}, hence drawing inference from this figure is difficult. No particular trait of cross was overrepresented at the extreme data points in Figure 3 (not shown). The modified EM algorithm fitted clearly only one component to the mixtures of normals, which implies that the dominance coefficients are normally distributed, with an estimated mean of 0.193 and estimated standard deviation of 0.312 (see Figure 4). The 95% confidence interval for the mean was −0.379 to 0.754 and for the standard deviation 0.230–0.375.

The modified EM algorithm suggested clearly fitting two components to the additive effects. The standard deviation of the first (second) component was 0.297 (0.509) with a 95% confidence interval of 0.281–0.329 (0.444–0.667). The mixing proportion for the first (second) component was 0.858 (0.142) with a standard error of 0.054. The distributions are shown in Figure 5. The combined distribution has somewhat thicker tails than a normal distribution. Hence, it seems that the majority of QTL show a small effect and only a few a larger effect. The mean of the absolute values of the combined distribution was 0.261 and was substantially lower compared to the mean of the absolute values of y_{add} (0.379, ±0.405, not shown elsewhere).

The truncation point for the fit of the exponential distribution to the additive effects was $c = 0.23$ units of phenotypic standard deviation. A number of 113 QTL with a smaller effect than this point were deleted. The fitted exponential distribution is shown in Figure 6. The rate parameter of the exponential distribution was $\hat{\lambda} = 5.81$ with a 95% confidence

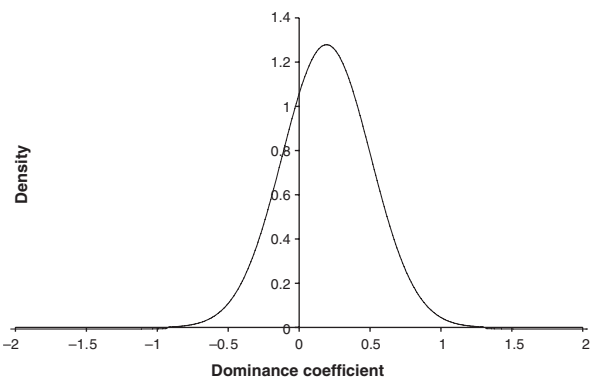

Figure 4 Fitted normal distribution to the dominance coefficients. The estimates mean was 0.193 (95% confidence interval −0.379 to 0.754) and estimated standard deviation was 0.312 with a 95% confidence interval of 0.230 to 0.375.

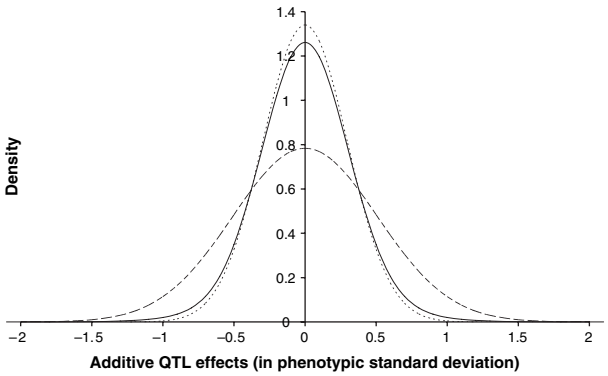

Figure 5 Fitted mixtures of normals to the additive quantitative trait loci (QTL) effects. Dotted line (dashed line) represents component one (component two) with mixing proportions 0.858 (0.142), and solid line the combined distribution. The standard deviation of component one (two) was 0.297 (0.509) with a 95% confidence interval of 0.281–0.329 (0.444–0.667).

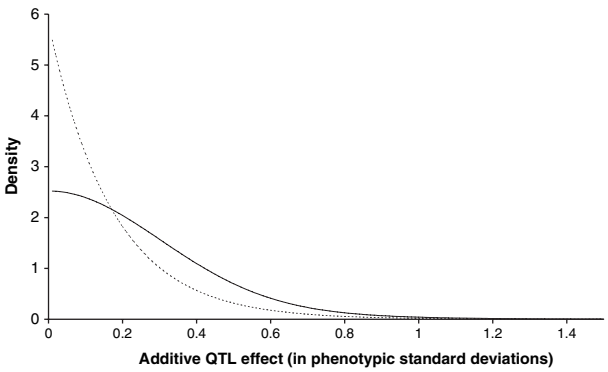

Figure 6 Fitted exponential distribution (dotted line) and combined mixture of normal distribution (solid line) to the additive quantitative trait loci (QTL) effects. The values of the combined mixture of normal distribution were multiplied by two, because only the positive part of the distribution is considered. The estimated mean of the exponential distribution was 0.172 with a 95% confidence interval of 0.143–0.196.

interval of 5.1–7.0. This resulted in a mean QTL effect of 0.172, which was lower compared to the mean of the combined distribution of the mixtures of normals.

Discussion

General assumptions

During the analysis some assumptions had to be made. It was assumed that each QTL is an independent repeated observation. This is a simplification because some QTL are found at very similar chromosomal positions in two crosses and hence they might be the same. Additionally, some traits are correlated and hence QTL for two correlated traits at a similar chromosomal position might be in fact a single pleiotropic QTL. However, estimates of QTL positions from a genome scan show in general large confidence intervals (Visscher *et al.* 1996) and hence are not precise enough to conclude that the QTL positions are the same. In general the inclusion of correlated records does not affect the distribution of the resulting data; however, it may imply that there are effectively fewer records than expected.

Weller *et al.* (2005) argued that it is problematic to derive distributions across traits and suggested to do this for each trait separately or at least for groups of traits with a similar heritability. In this study this would result in 34 different distributions, which would show large confidence intervals of their parameters due to limited number of QTL found for the single traits. Hence, the advantage of having many data points would be lost. Also it is unknown how good the distributions of QTL effects based on data points obtained from F_2 crosses can be generalized. Future results from genomewide QTL analysis using dense SNP markers in large data sets might allow detailed analysis of distribution of QTL for single traits of group of similar traits segregating within a population.

Distribution of QTL dominance coefficients

The distribution of dominance coefficients was largely unknown until now. The present study suggests a normal distribution of the dominance coefficients. Compared to the histogram of y_h (Figure 3) this distribution (Figure 4) shows a much lower density for extreme values, confirming the need to consider the heterogeneous error variances of the observations.

The phenomenon dominance has been discussed in the literature for several decades (Keightley 1996; Bourget 1999). Kacser & Burns (1973) developed a metabolic control theory, which modelled the phenotype as an end-product of enzyme activity. The enzyme activity causes a flux through metabolic pathways with a hyperbolic relationship. At a low (high) flux level an infinitesimal change of the enzyme activity results in a larger (smaller) change of the flux. This pattern is described by the control coefficient of the flux with respect to the enzyme activity. An important property of the control coefficients is that they sum up to one (Kacser & Burns 1973). Based on this model Kacser & Burns (1981) concluded that dominance will be more important for genes of large effects, and additionally dominance will be more often in the direction of the

genotype which produces a higher amount of end-product, i.e. $y_h > 0$ occurs more often than $y_h < 0$. Indeed the overall mean of the fitted distribution of the dominance coefficient is positive (Figure 4). To investigate the relationship between additive effects and dominance coefficients in more detail, y_h was plotted against $|y_{add}|$ (Figure 7). In order to minimize the sampling effects only $|y_{add}|$ being larger than three times their standard errors were considered. For low and intermediate values of $|y_{add}|$ (say below 0.4) all observed patterns of y_h seemed to be equally likely. For larger values over- and underdominance became less frequent and additionally there seems to be a trend towards recessivity. This is in agreement with observations made by Caballero & Keightley (1994) for genes affecting bristle characters in Drosophila. It does however not support the hypothesis that dominance is more important for genes of large effects. A possible explanation might be that the hyperbolic relationship between flux and enzyme activity does not hold for all traits considered (e.g. for meat quality traits).

Distribution of additive QTL effects

When using QTL mapping results to infer the distribution of the additive effects, the standard errors and the fact that only significant QTL are reported have to be taken into account (Hayes & Goddard 2001). When fitting mixtures of normals we did this by modifying the EM algorithm in order to account for the standard errors. Indeed, as becomes obvious from Equation (2), the number of components for the raw data was equal to the number of data points. However, we wanted to infer the components of the true effects, which was possible by the modifications made to the 'standard' EM algorithm. The fact that only significant QTL were considered was implicitly taken into account by augmenting the additive data set by QTL of opposite signs and setting the mean of the components equal to zero during the EM algorithm. The second way of modelling the distribution was the fit of an exponential distribution following Hayes & Goddard (2001). Both distributions revealed a leptokurtic distribution with many QTL of small effects and few of large effects. However, the exponential distribution was much more leptokurtic with a higher density for small effects (Figure 6). We did not test which distribution fit the data better, because we had only a truncated data set. A comparison test using this data set would favour a distribution that accounts less for the truncation. However, given the literature reports (Hayes & Goddard 2001; Orr 2003; Eyre-Walker & Keightley 2007) we believe that the mixtures of normals did not predict enough QTL with small effects.

In summary, the dominance coefficient followed a normal distribution with a positive mean. It seems that dominance coefficients are smaller for QTL with larger additive effects. A fit of mixtures of normals to truncated additive QTL effects revealed two components. Both the mixtures of normals and the exponential distribution were leptokurtic with many QTL of small and few of large effects. Compared to the exponential distribution and to literature reports, the mixtures of normals did not predict enough QTL with small effects. A formal comparison of both distributions was not possible due to the truncated data set. The derived distributions can be used until better data become available that allow a more detail look at single and trait specific QTL, for example, from genomewide association studies within populations using dense SNP data. These data could also be used for a formal comparison of different distributions.

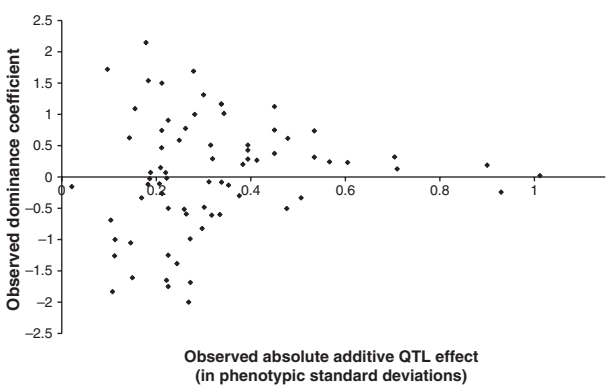

Figure 7 Observed dominance coefficients ($y_h = y_{dom}/|y_{add}|$) plotted against observed absolute additive quantitative trait loci (QTL) effects (n = 74).

Acknowledgements

J.B. was supported by a grant from the Deutsche Forschungsgemeinschaft, DFG. He thanks Teide-Jens Boysen (Institute of Animal Breeding and Husbandry, University of Kiel, Germany) and Heinz Bartenschlager (Institute of Animal Husbandry and Breeding, University Hohenheim, Germany) for their assistance in extracting the data from the publications.

References

Bourget D. (1999) The evolution of dominance. *Heredity*, **83**, 1–4.

Caballero A., Keightley P.D. (1994) A pleiotropic model of variation in quantitative traits. *Genetics*, **138**, 883–900.
Chamberlain A.J. (2004) Mapping Quantitative Trait Loci for Milk Production Traits in Dairy Cattle and Using Them in Marker Assisted Selection. PhD thesis, Institute of Land and Food Resources, University of Melbourne, Australia.
Dilda C.L., Mackay T.F.C. (2002) The genetic architecture of Drosophila sensory bristle number. *Genetics*, **162**, 1655–1674.
Eyre-Walker A., Keightley P.D. (2007) The distribution of fitness effects of new mutations. *Nat. Rev. Genet.*, **8**, 610–618.
Falconer D.S., Mackey T.F.C. (1996) Introduction to Quantitative Genetics. Longman, London.
Geldermann H.E., Müller G., Moser G., Reiner H., Bartenschlager S., Cepica A., Stratil J., Kuryl C., Moran R., Davoli R., Brunsch C. (2003) Genome-wide linkage and QTL mapping in porcine F_2 families generated from Pietrain, Meishan and Wild Boar crosses. *J. Anim. Breed. Genet.*, **120**, 363–393.
Goddard M.E. (2001) The validity of genetic models underlying quantitative traits. *Livest. Prod. Sci.*, **72**, 117–127.
Goddard M.E. (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*, **136**, 245–257.
Haley C.S., Knott S.A., Elsen J.M. (1994) Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics*, **136**, 1195–1207.
Hayes B., Goddard M.E. (2001) The distribution of the effects of genes affecting quantitative traits in livestock. *Genet. Sel. Evol.*, **33**, 209–229.
Kacser H., Burns J.A. (1973) The control of flux. *Symp. Soc. Exp. Biol.*, **27**, 65–104.
Kacser H., Burns J.A. (1981) The molecular basis of dominance. *Genetics*, **97**, 639–666.
Keightley P.D. (1994) The distribution of mutation effects on viability in Drosophila melanogaster. *Genetics*, **138**, 1315–1322.
Keightley P.D. (1996) A metabolic basis for dominance and recessivity. *Genetics*, **143**, 621–625.
Khatkar M.S., Thomson P.C., Tammen I., Raadsma H.W. (2004) Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol.*, **36**, 163–190.
Lynch M., Walsh B. (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associates, Inc. Publisher, Sunderland, MA. 980 pp.
McLachlan G.J. (1987) On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *J. R. Stat. Soc. C.*, **36**, 318–324.
McLachlan G.J., Peel D. (2000) Finite Mixture Models. Wiley Series in Probability and Statistics. John Wiley and Sons, New York.
Meuwissen T.H.E., Hayes B.J., Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, **157**, 1819–1829.
Müller E., Moser G., Bartenschlager H., Geldermann H. (2000) Trait values of growth, carcass and meat quality in Wild Boar, Meishan and Pietrain pigs as well as their crossbred. *J. Anim. Breed. Genet.*, **117**, 189–202.
Orr H.A. (2003) The distribution of fitness effects among beneficial mutations. *Genetics*, **163**, 1519–1526.
Otto S.P., Jones C.D. (2000) Detecting the undetected: estimating the total number of loci underlying a quantitative trait. *Genetics*, **156**, 2093–2107.
Peel D., McLachlan G.J. (1999) User's Guide to EMMIX, Version 1.3. University of Queensland, St. Lucia, Queensland, Australia.
Tanner M.A. (1993) Tools for Statistical Inference. Springer Verlag, New York.
Visscher P.M., Thompson R., Haley C.S. (1996) Confidence intervals in QTL mapping by bootstrapping. *Genetics*, **143**, 1013–1020.
Weller J.I., Shlezinger M., Ron M. (2005) Correcting for bias in estimation of quantitative trait loci effects. *Genet. Sel. Evol.*, **37**, 501–522.

Appendix A

The traits included and the number of quantitative trait loci (QTL) in each of the three crosses. European Wild Boar (W), Meishan (M) and Pietrain (P). For additional details regarding trait recording and definition see Müller *et al.* (2000) and Geldermann *et al.* (2003).

	Number of QTL observed in			
Trait	W × M	M × P	W × P	Total
Birth weight	1	2	–	3
21-day weight	4	–	–	4
35-day weight	2	1	–	3
Average daily gain	2	4	6	12
Food conversion ratio	2	2	1	5
Life weight at slaughter	7	3	6	16
Half carcass weight	3	1	4	8
Liver weight	4	–	–	4
Heart weight	2	2	5	9
Head weight	4	3	5	12
Carcass length	4	6	5	15
Bacon weight meat	4	5	6	15
Bacon weight external fat	2	3	4	9
Shoulder weight meat	3	2	6	11
Shoulder weight external fat	2	7	2	11
Chops weight meat	7	4	6	17
Average backfat depth	4	2	4	10
M. long. dorsi 13./14. rips backfat depth	1	3	1	5

Appendix A

(*Continued*)

Trait	Number of QTL observed in			
	W × M	M × P	W × P	Total
M. long. dorsi 13./14. rips meat area	1	6	4	11
M. long. dorsi 13./14. rips fat area	–	5	1	6
M. long. dorsi 13./14. rips meat: fat ration	4	6	2	12
Abdominal fat weight	1	5	3	9
Lean cuts	4	4	5	13
Fat cuts	3	6	4	13
M. long. dorsi 45 min pm pH	–	1	2	3
M. long. dorsi 45 min pm conductivity	1	1	3	5
M. long. dorsi 24 h pm conductivity	3	1	1	5
24 h pm colour	–	1	1	2
M. sem. 45 min pm pH	1	1	1	3
M. sem. 45 min pm conductivity	–	2	1	3
M. sem. 45 min pm stiffness	–	1	1	2
M. sem. 24 h pm pH	1	1	1	3
M. sem. 24 h pm conductivity	2	2	1	5
Stress resistance log CK20-value	–	4	2	6

M. long. dorsi, *Musculus longissimus* dorsi; pm, post mortem; M. sem., *Musculus semimembranosus*.

Appendix B

Using the same notation of the main text and let the observed QTL effect be y, the true effect be a and the error be e, i.e. $y_i = a_i + e_i$ (Equation 1 of the main text). Var(a_i) is the variance of the effects we wanted to infer (subsequently denoted by σ^2) and var(e_i) is the error variance (SE_i^2), which was assumed to be known from QTL mapping. According to mixed model theory a random QTL effect estimate can be written as

$$\begin{aligned}\hat{a}_i &= \left[SE_i^{-2} + \frac{1}{\sigma_j^2}\right]^{-1} * SE_i^{-2} y_i \\ &= \frac{SE_i^{-2}}{SE_i^{-2} + \sigma_j^{-2}} y_i \\ &= \frac{\sigma_j^2}{SE_i^2 + \sigma_j^2} y_i \\ &= h_i^2 y_i.\end{aligned} \qquad (A1)$$

The expectation of the sums of squares of a_i belonging to the same component j given y_i is

$$E\left(\sum_{i=1}^{n} z_{ij} a_i^2 | y_i\right) = \sum_{i=1}^{n} z_{ij}(\hat{a}_i^2 + PEV_i),$$

where z_{ij} is the indicator variable indicating to which component i belongs. PEV_i denotes for the predicted error variance and is

$$\begin{aligned}PEV_i &= \left[SE_i^{-2} + \sigma_j^{-2}\right]^{-1} \\ &= \frac{SE_i^2 \sigma_j^2}{SE_i^2 + \sigma_j^2} \\ &= h_i^2 SE_i^2.\end{aligned}$$

Now,

$$\sum_{i=1}^{n} z_{ij}\sigma_j^2 = \sum_{i=1}^{n} z_{ij}(\hat{a}_i^2 + h_i^2 SE_i^2), \text{ and}$$

$$\sigma_j^2 = \frac{\sum_{i=1}^{n} z_{ij}(\hat{a}_i^2 + h_i^2 SE_i^2)}{\sum_{i=1}^{n} z_{ij}},$$

The indicator z_{ij} is unknown and has to be replaced by its conditional expectation (see Equation (3) of the main text), resulting in

$$\sigma_j^2 = \frac{\sum_{i=1}^{n} \tau_{ij}(\hat{a}_i^2 + h_i^2 SE_i^2)}{\sum_{i=1}^{n} \tau_{ij}}. \qquad (A2)$$

Substituting (A1) into (A2) yields

$$\sigma_j^2 = \frac{\sum_{i=1}^{n} \tau_{ij} h_i^2 (h_i^2 y_i^2 + SE_i^2)}{\sum_{i=1}^{n} \tau_{ij}},$$

which is equivalent to Equation (5) of the main text.