Genet. Sel. Evol. 37 (2005) 635–655

DOI: 10.1051/gse:2005019

Original article

Efficiency of population structures for mapping of Mendelian and imprinted quantitative trait loci in outbred pigs using variance component methods

Henri C.M. HEUVEN*, Henk BOVENHUIS, Luc L.G. JANSS, Johan A.M. VAN ARENDONK

Animal Breeding and Genetics group, Animal Sciences Group, Wageningen University and Research Centre, P.O. Box 338, 6700AH Wageningen, The Netherlands

(Received 4 October 2004; accepted 25 June 2005)

Abstract – In a simulation study different designs for a pure line pig population were compared for efficiency of mapping QTL using the variance component method. Phenotypes affected by a Mendelian QTL, a paternally expressed QTL, a maternally expressed QTL or by a QTL without an effect were simulated. In all alternative designs 960 progeny were phenotyped. Given the limited number of animals there is an optimum between the number of families and the family size. Estimation of Mendelian and parentally expressed QTL is more efficient in a design with large family sizes. Too small a number of sires should be avoided to minimize chances of sires to be non-segregating. When a large number of families is used, the number of haplotypes increases which reduces the accuracy of estimating the QTL effect and thereby reduces the power to show a significant QTL and to correctly position the QTL. Dense maps allow for smaller family size due to exploitation of LD-information. Given the different possible modes of inheritance of the QTL using 8 to16 boars, two litters per dam was optimal with respect to determining significance and correct location of the QTL for a data set consisting of 960 progeny. The variance component method combining linkage disequilibrium and linkage analysis seems to be an appropriate choice to analyze data sets which vary in marker density and which contain complex family structures.

imprinting / quantitative trait loci / simulation / pig / designs

1. INTRODUCTION

In the last decade several genome scans have revealed quantitative trait loci for numerous traits in pigs [2, 4–7, 10, 15, 16, 21, 23, 24, 26]. Mendelian as well as imprinted modes of inheritance were observed [4, 26]. The determination of

* Corresponding author: henri.heuven@wur.nl

the causal mutation and the mode of inheritance have been reported for only a few QTL in pigs, *e.g.* a paternally expressed mutation in an intron of the *IGF2*-gene affects muscle growth, fat deposition and heart size [27]. Milan *et al.* [20] identified a possible causative mutation in the PRKAG3-gene and its effect on glycogen content in pig skeletal muscle. Since most of the pig breeding programs rely on cross breeding, the mode of inheritance can be an important factor for successfully applying marker assisted selection (MAS) with regards to QTL.

In humans it was shown that more than 50% of the genes showed preferential expression of the paternal or maternal allele [25]. Given the homology between humans and pigs this could also be the case for pigs. In order to utilize QTL for MAS the confidence regions surrounding the QTL have to be reduced and the mode of inheritance has to be determined. Theory and methods to map the position of QTL have been proposed [3, 17, 18]. The variance component (VC) method that combines linkage (LA) and linkage disequilibrium (LD) between markers and QTL is gaining considerable attention [8, 11]. VC methods allow for simultaneous estimation of systematic, polygenic and QTL (haplotype) effects while accounting for more complex data structures and pedigrees, *e.g.* designs with full sibs nested within half sibs or with mixed paternal and maternal half sibs. Additionally different genetic models can easily be compared using VC methods. Small effective population sizes generate LD among parental haplotypes which can be combined with the traditional LA in the VC method as described by Meuwissen *et al.* [19].

Most pig studies were performed in crosses of divergent lines, whereas interest is now growing to validate and apply QTL mapping within commercial pig populations. Commercial pig populations exhibit relatively complex pedigree structures with relatively small (full sib) family sizes. The application of VC methods would be ideal for this type of structure, but properties of VC methods for QTL mapping have not been widely studied, VC methods have not been optimized for experimental designs of pig populations and VC methods have not been considered for mapping of imprinted QTL. Several factors can be varied while setting up mapping studies for pigs, *e.g.* the number of markers, marker spacing, the number of parents and the number of genotyped and phenotyped offspring. Usually there is a trade-off between the number of animals and the number of markers to be genotyped to keep the genotyping cost within limits. Changing the population structure will influence the ability to identify significant QTL [1, 13, 14].

The aim of this study was to investigate the efficiency of various experimental designs for mapping of QTL with different modes of inheritance

(Mendelian, paternal and maternal imprinting) in pigs with Variance Component methods using simulation.

2. MATERIALS AND METHODS

2.1. Data simulation

2.1.1. Simulation of pedigrees and traits

The pedigrees used in the simulation model were set up into two parts. In the first part, the population was simulated by random mating animals in 100 successive generations while in the second part different mapping populations were generated as described below. In the first part, 120 males and 120 females were used in each generation to produce the next generation. Each mating resulted in one male and one female offspring. Each male and female was used once as the parent of the next generation.

Genotypes were generated independently for eight markers, starting with six equally frequent alleles in the base generation, and one QTL on a single chromosome. A unique QTL allele was assigned to each marker haplotype in the base generation. The map distance between markers was either 2 or 10 cM. The QTL was simulated half way between marker three and marker four. Its map distance to either marker was therefore 1 or 5 cM for a marker distance of 2 or 10 cM, respectively. In each generation for each individual its paternal (maternal) haplotype was determined from the haplotypes of the sire (dam). Each haplotype was built up by randomly choosing one of the parental haplotypes. The alleles at each position were a copy of the parental haplotype taking crossing over into account. A Poisson distribution, using the distance to the next marker or QTL as parameter, was used to determine the probability of an uneven number of cross-overs.

In the final generation the QTL allele with the frequency closest to 0.2 was assigned a favorable effect on the phenotype such that an approximately constant QTL variance was generated. A null value was assigned to all other alleles. The variance due to the QTL was calculated as $Vqtl = 2p(1 - p)\alpha^2$ for the Mendelian inherited trait and as $Vqtl = 4p(1 - p)\alpha^2$ for the parentally imprinted traits [6], where p is the observed frequency of the favorable QTL allele in the (final) offspring generation and α *is* the gene substitution effect. By varying α, the *Vqtl* was fixed at 10% of the total phenotypic variation, which was set to be 1.0, *i.e.* a heritability of 0.1 for the QTL effect.

Polygenic effects were simulated for base animals and subsequently for each animal in consecutive generations based on the average breeding value

Table I. Alternative family structures considered for mapping populations consisting of 960 progeny.

1 litter/sow				2 litters/sow			
Boars	Sows	PHS	MH	Boars	Sows	PHS	MHS
4	96	240	10	8	48	120	20
8	96	120	10	16	48	60	20
24	96	40	10	48	48	20	20

PHS = paternal half-sibs; MHS = maternal half-sibs.

of the parents and a Mendelian sampling term. These terms were drawn from $N(0,0.2)$, *i.e.* the heritability of the polygenic effect was set at 0.2.

For each individual, four different trait phenotypes were generated: P_{mend}, P_{pat}, P_{mat} and P_{noQ}. P_{mend} is a trait with a Mendelian additive QTL-effect. P_{pat} and P_{mat} are traits with respectively a paternally and maternally expressed QTL-effect. No QTL was simulated for P_{noQ}. Phenotypes were generated according to the following model:

$$y = \mu + \textit{polygenic effect} + \textit{QTL effect} + e \quad (1)$$

where e is sampled from $N(0,\sigma_e^2)$. For P_{mend}, P_{pat}, and P_{mat} $\sigma_e^2 = 0.7$ and for P_{noQ} $\sigma_e^2 = 0.8$. Phenotypes were generated for offspring, *i.e.* the mapping population, only.

2.1.2. Mapping population structures

Six different family structures were generated for the mapping population for each population generated in the first part. The six populations differed with respect to the number of sires, dams and offspring as is shown in Table I.

In total 100 replicates were simulated for each family structure. Starting from 120 males and 120 females allowed parents for the six mapping populations to be chosen such that they were neither full-sibs nor half-sibs. The six mapping populations were generated such that the total number of progeny equaled 960. Different boars were used to produce the two litters per sow. Phenotypic values were assumed not to be dependent on the parity of the sow.

2.2. Analysis of simulated data

2.2.1. Statistical models

Simulated data was analyzed using the following two models:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wv} + \mathbf{e} \tag{2}$$

and

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{W_s v_s} + \mathbf{W_d v_d} + \mathbf{e} \tag{3}$$

where **y** is a vector of N observations, **b** is a vector of fixed effect (in the present study only the mean), **u** is a vector of random effects due to unlinked genes for each animal (polygenic effect), **v** is a vector of random effects due to the QTL (haplotype effect) and **e** are residuals. In model (2) **v** contains a maternal and a paternal haplotype effect where each record is linked to a set of two haplotype effects. In model (3) a separate random effect is included for the maternal ($\mathbf{v_d}$) and paternal ($\mathbf{v_s}$) haplotypes to allow for differences in size of effects (allow for imprinting). The random effects (**u**, **v**, $\mathbf{v_s}$, $\mathbf{v_d}$, **e**) are assumed to be normally distributed with mean zero and variance $\sigma^2_{\mathbf{u}}$, $\sigma^2_{\mathbf{v}}$, $\sigma^2_{\mathbf{vs}}$, $\sigma^2_{\mathbf{vd}}$ and $\sigma^2_{\mathbf{e}}$. **X**, **Z**, **W**, $\mathbf{W_s}$ and $\mathbf{W_d}$ are design matrices for the effects of **b**, **u**, **v**, $\mathbf{v_s}$, and $\mathbf{v_d}$ respectively.

The phenotypic variance of the observations using model (2) is:

$$\mathbf{V} = \mathbf{ZAZ}'\sigma^2_{\mathrm{u}} + \mathbf{WG_pW}'\sigma^2_{\mathrm{v}} + \mathbf{R} \tag{4}$$

and

$$\mathbf{V} = \mathbf{ZAZ}'\sigma^2_{\mathrm{u}} + \mathbf{W_sG_pW}'_{\mathbf{s}}\sigma^2_{\mathrm{vs}} + \mathbf{W_dG_pW}'_{\mathbf{d}}\sigma^2_{\mathrm{vd}} + \mathbf{R} \tag{5}$$

for model (3), where **A** is the numerator relationship matrix based on additive genetic relationships including five generations prior to the parent generation, $\mathbf{G_p}$ is the matrix containing the IBD probabilities of a putative QTL at location **p** and $\mathbf{R = I}\,\sigma^2_{\mathbf{e}}$ (**I** is an identity matrix).

The linkage disequilibrium information is included by calculating separately for each pair of parental haplotypes their IBD probability taking marker information and five generations of known pedigree into account.

Here we assume that marker data is available for parents and offspring only and that their phases are known. The linkage information is based on the IBD probability between a parental and an offspring haplotype given that both animals are genotyped. The method used to estimate the $\mathbf{G_p}$ matrix is described in detail by Meuwissen [17, 18]. The subroutines provided were combined in a package called LDLA [12].

Since the IBD probabilities are calculated for each pair of haplotypes at the time, the overall matrix can become non-positive definite. This situation occurs especially for a large number of generations and/or a small effective population size and/or dense marker maps. The number of generations was taken as 100 and the effective population size was taken as 240. Eigenvalues were calculated for the LD-part of each $\mathbf{G_p}$ matrix and when the smallest one was less than zero, all eigenvalues were bent towards their mean such that the smallest eigenvalue attained a value of 0.01.

Ignoring the covariance between the parental haplotypes results in an analysis based on linkage information only, *i.e.* LA-only. A comparison of the results of LDLA-analyses and LA-only analyses show the added value of including linkage disequilibrium information.

2.2.2. Test statistics

Assuming multivariate normality of the data $\sim N(Xb,V)$ the ASReml-package [9] was used to calculate the maximum log likelihood and the variance components for each bracket (**p**) using the appropriate $\mathbf{G_p}$. A likelihood ratio test (LRT) was calculated as follows:

$$\text{LRT} = -2 * (\text{log likelihood } (H_1) - \text{log likelihood } (H_0)).$$

Where log likelihood (H_1) is the likelihood for a model with a QTL-effect and it is calculated for each bracket. Log likelihood (H_0) is based on a model excluding the QTL-effect(s).The LRT-statistic has a Chi-square distribution with either one or two degrees of freedom depending upon the number of QTL parameters that were constrained, *i.e.* model (2) or model (3), respectively. The LRT statistic does not take into account that multiple tests are performed along the chromosome, but our simulations with no-QTL data (P_{noQ}) will provide an estimate of chromosome-wise false positive rate.

3. RESULTS

3.1. Power

The percentage of simulations with a significant QTL (p-values < 0.05 or 0.01) for P_{noQ}, *i.e.* false positives, is given in Table II. At a significance level of 0.05 five out of 100 simulations are expected to show a significant QTL. In almost all situations explored, this is exceeded. At 0.01 the observed number is closer to the expected number. Fewer false positives were found for

Table II. Percentage of false positive simulations, based on 100 replicates, for a phenotype without a QTL effect (P_{noQ}) for different types of analysis, models, map distances and significance levels per population type.

Type of analysis:			LDLA[1]				LA-only			
		Model[2]:	2		3		2		3	
# of sires	# of dams	# of progeny	0.05[3]	0.01	0.05	0.01	0.05	0.01	0.05	0.01
			map distance 2 cM							
4	24	10	5	3	5	3	4	1	4	2
8	12		10	1	7	1	4	0	6	0
24	4		3	1	3	0	7	0	7	0
8	12	20	5	3	6	3	7	0	6	2
16	6		2	0	4	2	6	1	7	1
48	2		9	0	10	0	5	3	5	2
			map distance 10 cM							
4	24	10	6	1	8	0	8	1	10	3
8	12		9	1	6	0	9	1	8	1
24	4		8	3	6	1	10	5	9	4
8	12	20	8	1	7	1	7	2	5	1
16	6		9	2	9	0	12	1	9	0
48	2		12	3	6	0	9	3	7	1

[1] LDLA analysis combines LD- and LA-information; LA-only ignores LD-information.
[2] In model (2) a single variance component is fitted for paternal and maternal haplotypes, in model (3) a separate component is fitted for each. These models were tested against a model without a QTL-effect.
[3] Nominal significance level.

the 2 cM map when compared to the 10 cM map. Model (3) requires a more extreme test statistic to reach the significance threshold, *i.e.* two QTL components are estimated, but the number of false positives is only slightly lower when compared to model (2).

For P_{mend}, P_{pat} and P_{mat} the percentage of simulations where the test statistic was significant (p-value < 0.01) is given in Table III and partly in Figure 1. Substantial differences can be observed for each of the three modes of inheritance. The power for a maternally expressed QTL, *i.e.* P_{mat}, is much smaller than what was obtained for P_{mend} and P_{pat}. A significant QTL is found in 44 to 96%, 46 to 91% and 10 to 90% of the simulations at a significance level of 0.01 for P_{mend}, P_{pat} and P_{mat}, respectively.

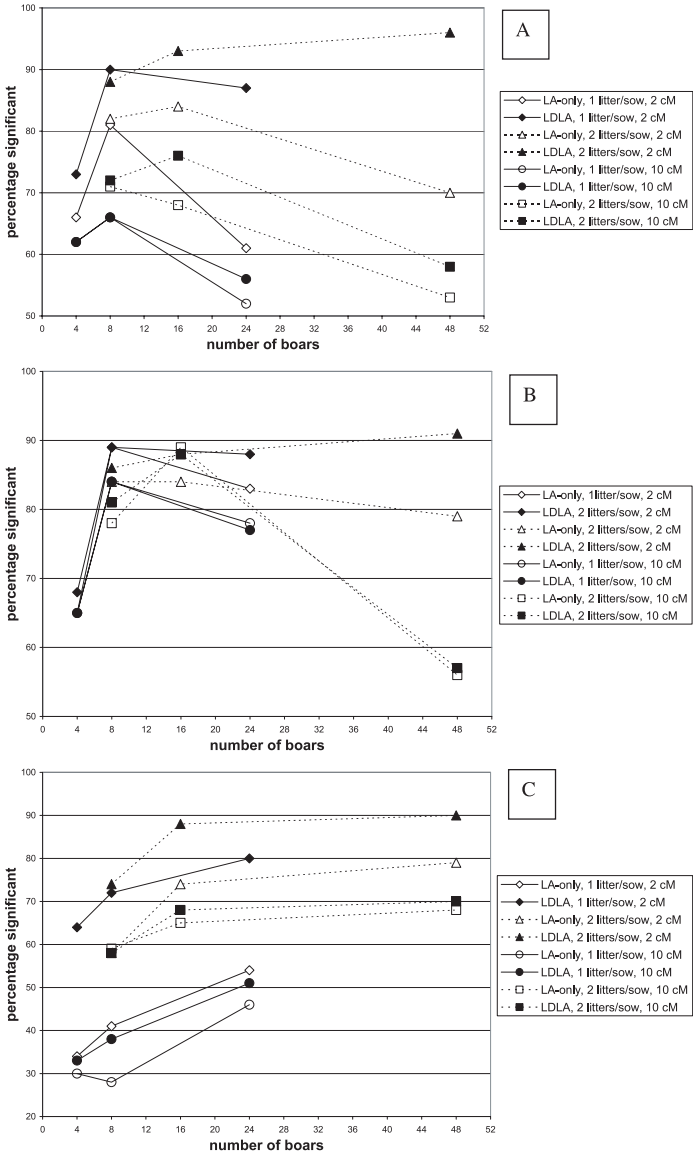

Figure 1. Percentage of simulations with a significant QTL ($p < 0.01$) for phenotypes affected by a Mendelian QTL (P_{mend}, A), a paternal expressed QTL (P_{pat}, B) and maternal expressed QTL (P_{mat}, C). Separate lines are given for analyses based on combined linkage and linkage disequilibrium information (LDLA) and for analyses ignoring LD information (LA-only), 2 and 10 cM map distance and for 1 and 2 litters per dam.

Table III. Percentage of simulations showing a significant QTL effect at a p-value of 0.01, per trait for different types of analysis and different models.

Type of analysis:			LDLA[1]				LA-only			
Model[2]:			2		3		2		3	
# of sires	# of dams	# of progeny	2 cM[4]	10 cM	2 cM	10 cM	2 cM	10 cM	2 cM	10 cM
						$\mathbf{P_{mend}}$[3]				
4	96	10	73	62	73	60	66	62	66	58
8	96		90	66	86	63	81	66	77	60
24	96		87	56	70	48	61	52	55	47
8	48	20	88	72	82	66	82	71	79	64
16	48		93	76	83	63	84	68	75	63
48	48		96	58	85	50	70	53	61	44
						$\mathbf{P_{pat}}$				
4	96	10	66	65	68	65	65	65	65	65
8	96		89	84	89	84	89	84	89	84
24	96		84	73	88	77	82	73	83	78
8	48	20	83	80	86	81	85	79	84	78
16	48		81	80	88	88	87	81	84	89
48	48		79	48	91	57	74	46	79	56
						$\mathbf{P_{mat}}$				
4	96	10	49	15	64	33	27	10	34	30
8	96		62	20	72	38	23	16	41	28
24	96		52	24	80	51	21	17	54	46
8	48	20	62	42	74	58	38	41	58	59
16	48		68	47	88	68	59	45	74	65
48	48		75	52	90	70	66	50	79	68

[1] LDLA analysis combines LD- and LA-information; LA-only ignores LD-information.

[2] In model (2) a single variance component is fitted for paternal and maternal haplotypes, in model (3) a separate component is fitted for each.

[3] P_{mend}, P_{pat} and P_{mat} are phenotypes affected by a Mendelian, paternal and maternal expressed QTL, respectively.

[4] Map distance among the markers.

Power is considerably higher for the 2 cM map as compared to the 10 cM map. A denser map apparently allows for a better discrimination between the presence and absence of a segregating QTL.

The additional value of LD-information is evident for all populations using a 2 cM map except for P_{pat} where including LD resulted in similar power as

was obtained using linkage mapping (LA-only). In the latter case the power is already high for P_{pat} due to the relatively large paternal half sib groups. LD information is therefore of added value for the population based on 48 sires.

Allowing for independent variance components for paternal and maternal haplotype effects, *i.e.* model (3), resulted in an increased power for P_{mat} and similar results for P_{pat}, compared to a single haplotype component model (model 2). The appropriate model for P_{mend}, model (2), showed a higher power than what was obtained under model (3) probably due to the lower significance threshold required as only one component is fitted and because segregation from both the dams and the sires can be utilized.

Clear differences in power were also observed for different population structures. There is an increase in power going from 4 to 8 sires and little additional power is gained when the number of sires exceeds 16. It is even declining for P_{mend} and P_{pat}. Using few sires increases the risk that none of them is segregating while using many sires with a few offspring each, decreases the ability to estimate QTL effects and thus the power to detect a significant QTL. However, in case the linkage disequilibrium information increases, information across families is combined which decreases the importance of family size.

More progeny per dam and half sib groups of 60 to 120, *i.e.* 16 and 8 sires respectively, gave the highest power, indicating that there should be a sufficient number of parents segregating for the QTL and the progeny groups should be of moderate size to allow reliable estimation of QTL effects. Either a low number of parents or a small number of progeny per family reduced the power. For the detection of maternally expressed QTL in pigs, maternal progeny groups should be as large as possible. For P_{mat} the advantage of a larger maternal half sib group is more pronounced compared to P_{mend} and P_{pat}. However, maternal progeny group size is limited in pigs. The number of dams that will be used in QTL experiments is usually sufficient to find several dams that are segregating for the QTL.

3.2. Location

The estimation of the correct location of the QTL is very important for the design of subsequent fine mapping experiments. In Table IV the percentage of simulations is given where the maximum likelihood ratio statistic was between the 2nd and the 5th marker, *i.e.* the correct bracket plus and minus one. The interval covered is in this case 6 and 30 cM for a map distance between markers of 2 and 10 cM, respectively. The results for P_{mend} based on model (2) and the results based on model (3) are shown for P_{pat} and P_{mat}, respectively.

Table IV. Percentage of the simulations with the maximal likelihood ratio test statistic between the 2nd and 5th marker at three significance levels per trait for each population, type of analyses and map distance. Model (2) was applied for P_{mend} and model (3) for P_{pat} and P_{mat}.

Type of analysis:			LDLA[1]						LA-only					
			P_{mend}[2]											
# of	# of	# of	2 cM[3]			10 cM			2 cM			10 cM		
sires	dams	progeny	all	0.05[4]	0.01	all	0.05	0.01	all	0.05	0.01	all	0.05	0.01
4	96	10	79	86	88	77	92	94	72	76	77	80	94	95
8	96		88	89	91	87	93	95	68	71	74	86	90	97
24	96		81	83	85	86	89	88	71	72	75	83	85	87
8	48	20	80	81	83	88	93	96	67	72	76	90	93	96
16	48		94	94	95	91	93	93	71	70	74	89	91	94
48	48		97	97	97	78	82	79	69	72	74	74	76	79
			P_{pat}											
4	96	10	76	90	91	76	93	95	71	87	88	78	91	95
8	96		78	81	82	92	94	96	77	79	80	90	93	95
24	96		83	84	84	91	93	94	72	72	77	91	95	95
8	48	20	78	84	85	89	96	96	73	81	81	89	95	97
16	48		85	87	89	92	95	97	82	82	85	88	91	92
48	48		94	96	96	80	87	86	71	73	75	77	82	80
			P_{mat}											
4	96	10	86	95	95	69	80	88	68	67	71	63	78	83
8	96		88	89	90	77	90	89	62	74	76	76	78	82
24	96		79	83	85	87	88	88	66	69	69	84	86	89
8	48	20	86	88	93	80	84	88	69	69	74	77	84	83
16	48		86	88	89	89	90	91	69	71	73	89	94	94
48	48		91	91	90	87	91	91	72	74	73	85	90	91

[1] LDLA analysis combines LD- and LA-information; LA-only ignores LD-information.
[2] P_{mend}, P_{pat} and P_{mat} are phenotypes affected by a Mendelian, paternal and maternal expressed QTL, respectively.
[3] Map distance between markers.
[4] Nominal significance level; all = averaged over all simulations.

The percentage of simulations with a significant QTL segregating and with a correct location, *i.e.* between markers 2 and 5, ranges from 62 to 97. It indicates that there is a real probability that an erroneous conclusion can be drawn about the location. Using a stringent threshold, *e.g.* 0.01, results in only slightly more

simulations having a maximum at the correct location compared to the 0.05 threshold.

The results from the 2 cM map can not be compared with the 10 cM map because the interval differs. LD information has a large effect on the correct positioning of the QTL for the 2 cM map but only a minor effect for the 10 cM map.

The populations sired by eight or sixteen sires have the highest scores for correct location of the QTL and additional sires do not add much or even have a lower percentage of correct locations, *e.g.* P_{mend} and P_{pat} for populations with 24 or 48 sires, similarly to what was observed for power. For P_{mat} not only the power to indicate a significant QTL was lower for all populations when compared to P_{mend} and P_{pat} but also the location is estimated less accurate. For P_{mat} only minor differences were observed among population types.

3.3. Estimated components

In Table V estimated components are expressed as a percentage of the simulated components and averaged over all simulations. QTL and polygenic variance components are expressed as heritabilities. The components shown are from model (2) for P_{mend} and from model (3) for P_{pat} and P_{mat}.

Variance components estimated using the model without QTL, *i.e.* assuming the null hypothesis, were identical for the 2 cM and 10 cM map because the same 100 random seeds were used for each simulation. For the model without a QTL-effect the QTL-variance is picked up by polygenic component which therefore reflects the total genetic variance.

For models with a QTL effect, the polygenic component is underestimated to some extent for all the analyses. However, only seven generations were included for the calculation of the numerator relationship matrix (A). Therefore the additive genetic variance in generation 95 instead of the base generation was estimated.

In most cases also the estimate of the QTL variance is higher by 10–50% if linkage disequilibrium is accounted for (LDLA analyses). When averages are calculated for significant simulations only, the polygenic component is in all cases even more underestimated, and the QTL component is increased (data not shown).

Using two components to fit the QTL variance, *i.e.* model (3), instead of one results in similar estimates of the QTL variance for P_{mend} (data not shown). For P_{mend} the paternal QTL component (h^2_{vs}) is slightly larger than the maternal component (h^2_{vd}) when few sires are used and the reverse occurs if many sires are used. Fitting model (3) in this case indicated that a QTL is differently

Table V. Average phenotypic variances and heritabilities as a percentage of the simulated values for each trait type of analysis, map distance and population type.

			Type of analysis:		LDLA[1]						LA-only					
			Map distance		2 cM[3]			10 cM			2 cM			10 cM		
# of	# of	# of	under H_0:													
										$\mathbf{P_{mend}}$[2]					
sires	dams	progeny	σ_p^2	h_a^2	σ_p^2	h_a^2	h_v^2	σ_p^2	h_a^2	h_v^2	σ_p^2	h_a^2	h_v^2	σ_p^2	h_a^2	h_v^2
4	96	10	101	137	105	91	122	102	87	109	101	95	88	101	82	106
8	96		100	131	104	84	127	102	78	123	100	89	90	100	79	110
24	96		101	136	105	87	133	102	87	113	101	92	97	101	88	103
8	48	20	101	139	105	92	126	103	87	120	101	98	88	101	88	105
16	48		101	138	105	87	138	102	86	121	101	94	95	101	88	107
48	48		101	133	105	88	129	102	81	128	100	91	94	100	83	114
										$\mathbf{P_{pat}}$					
4	96	10	96	118	106	82	131	104	90	107	101	85	103	101	85	100
8	96		96	118	105	76	135	103	81	118	99	80	101	100	81	104
24	96		99	129	104	80	138	102	79	122	100	83	104	100	81	108
8	48	20	98	122	104	86	123	103	85	115	100	90	89	100	83	104
16	48		99	130	104	84	132	101	85	114	100	88	97	100	88	102
48	48		101	136	104	84	129	101	77	124	100	89	96	100	80	114
										$\mathbf{P_{mat}}$					
4	96	10	106	156	106	85	157	103	80	141	102	82	123	101	77	134
8	96		104	146	105	87	137	102	79	130	101	90	104	101	84	113
24	96		103	143	105	80	151	102	82	124	100	83	112	100	81	115
8	48	20	105	156	105	86	149	102	89	126	101	93	104	101	89	113
16	48		103	147	105	81	152	103	83	131	101	88	107	101	81	119
48	48		101	132	105	76	152	102	78	132	100	80	109	100	80	120
										$\mathbf{P_{noQ}}$					
4	96	10	103	96	103	87		104	89		103	84		104	86	
8	96		101	91	102	79		102	82		101	79		102	81	
24	96		100	89	102	74		101	82		100	74		101	78	
8	48	20	101	87	101	75		101	80		101	77		101	79	
16	48		100	91	102	78		101	85		100	79		101	85	
48	48		101	90	103	74		101	82		101	75		100	79	

[1] LDLA analysis combines LD- and LA-information; LA-only ignores LD-information.

[2] P_{mend}, P_{pat} and P_{mat} are phenotypes affected by a Mendelian, paternal and maternal expressed QTL, respectively. P_{noQ} is phenotype without QTL effect. For P_{mend} results are shown based on model (2) while for P_{pat} and P_{mat} results based on model (3) are shown.

[3] Map distance between markers.

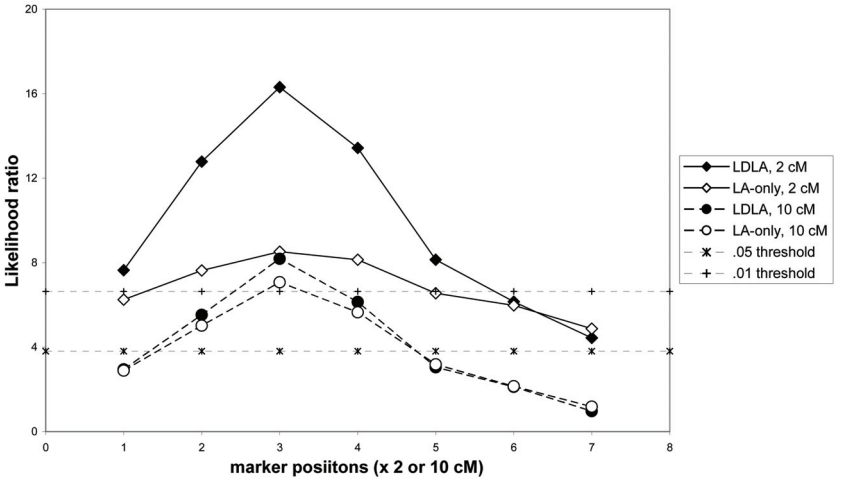

Figure 2. Average likelihood ratio statistics across the chromosome for LDLA and LA-only analysis and for 2 and 10 cM map distance for phenotypes affected by a Mendelian expressed QTL (P_{mend}). The population based on 16 sires is shown.

expressed among the paternal and maternal haplotypes. When this is observed in real data the conclusion could be that maternal and paternal haplotypes are differentially expressed where it could be the result of the specific population structure. In cases of maternal or paternal expression, *i.e.* P_{pat} and P_{mat}, both the maternal and paternal components are fairly well estimated using model (3).

For each bracket the likelihood ratio test statistic, averaged over all simulations, for P_{mend} is shown in Figure 2 for populations with 20 progeny per dam sired by 16 boars.

The shapes of the curves for P_{pat}, P_{mat} as well as for other populations are very similar in shape but different in magnitude. The shape is fairly symmetric around position three indicating that the larger number of brackets to the right of the QTL does not add additional information to estimate the variance component. Comparing the 2cM and 10 cM map results indicate that the LRT covers approximately the same distance above the 0.01 threshold, *i.e.* ca 10 cM for LDLA and 7 cM for LA-only. This might indicate that the confidence region for the location of the QTL is not much affected by the map distance among markers. The additional value of LD information is obvious at the peak but also at position seven. At this position no QTL should be detected and due to including LD-information the LRT-statistic is lower for LDLA compared to LA-only. A lower LRT-statistic was also observed for P_{noQ} when comparing

LDLA with LA-only, *i.e.* fewer false positives for LDLA were observed when compared to LA-only (Tab. I).

4. DISCUSSION

4.1. Choice of experimental design for pig QTL studies and accuracies

The present study proposes a design of family structure in a pig population that could be applied to map a QTL with different modes of inheritance for two different map distances among markers. Experiments to detect QTL (and their mode of inheritance) will be restricted in size in many cases because marker assisted selection has added value in current breeding programs for traits that can be measured late in the production cycle, *e.g.* meat quality, or that can be measured on one sex only or that are very expensive to measure, *e.g.* disease resistance. The number of animals on which the phenotypes will or can be collected will therefore be limited and therefore the number of animals to be genotyped for QTL discovery will also be fixed. Given a fixed number of animals the results in Tables III and IV indicate that the number of haplotypes and the accuracy with which each haplotype can be estimated should be balanced. This balance, however, differs depending on the mode of inheritance. In this study the frequency of the favorable QTL-allele was chosen to be around 0.2. In case the QTL-allele is rare more parents should be used to increase the chance of including segregating families. The results by Lee and van der Werf [13] based on full sib or half sib pedigrees also indicated that the number of families should be balanced with family size for the highest accuracy. Van der Beek *et al.* [1] used a deterministic approach to estimate the power of different designs. In Table VI the results obtained in this study using LA-only analysis for P_{mend} are compared with the deterministically determined values using a heritability of 0.2, a QTL-effect of 0.56 and an allele frequency of 0.2, *i.e.* the values used in the simulations. It should be noted that the deterministic approach does not fully account for the actual population structure. It considers either unrelated paternal half sib groups or unrelated full sib groups.

The power estimated by simulation is less than the power determined deterministically for populations with 10 offspring per dam, *i.e.* 1 litter/sow, and it is higher for populations with 20 offspring per dam using half sib formulas. The reverse occurred when full sib formulas were used. However, the ranking of populations with regards to power was fairly similar. Deterministic formulae provided by van der Beek *et al.* [1] are therefore useful for a quick comparison of different experimental designs but only apply for the segregation of a Mendelian QTL while ignoring LD information.

Table VI. Deterministic and simulated power for a phenotype affected by a Mendelian segregating QTL (P_{mend}) for each map distance and population type using linkage information only and a single variance component model (model 2).

		2 cM		10 cM	
# sires	Half-Sibs	determ.[1]	simulated	determ.	simulated
4	240	0.73	0.66	0.71	0.62
8	120	0.75	0.81	0.71	0.66
24	40	0.61	0.61	0.56	0.52
8	120	0.75	0.82	0.71	0.71
16	60	0.69	0.84	0.64	0.68
48	20	0.43	0.70	0.38	0.53
# dams	Full-Sibs				
96	10	0.52	0.61	0.45	0.52
48	20	0.76	0.70	0.70	0.53

[1] Deterministic calculations based on formulæ given by van der Beek *et al.* [1].

Including more litters per dam seem to be advantageous assuming that the phenotype of the offspring is not affected by the parity of the dam. Alternatively, parity could be corrected for. Using more litters per dam decreases the number of maternal haplotypes and increases the number of observations per haplotype. It therefore increases the ability to estimate its effect more accurately. Using different boars to sire the different litters allows for including a maternal genetic component in the model which would otherwise be confounded. Using different sires for subsequent litters also increases the ability to determine the phase of the markers and therefore the construction of haplotypes. In this study, however, this aspect was not considered because linkage phases were assumed to be known.

The objective of genome scans is not only to discover significant QTL but also to locate the position as accurately as possible. Especially for a dense map linkage disequilibrium is of added value to estimate the correct location. In general it seems that populations based on many sires (> 16) and therefore smaller families are less suited to estimate the correct location.

4.2. Significance level and hypothesis testing

In QTL studies, there is a serious risk of obtaining false positive QTL by chance due to the large number of tests as well as the large number of traits that are analyzed. Detection of spurious QTL was demonstrated in this study by

showing significant QTL for the trait P_{noQ}, *i.e.* a trait for which no QTL effect was simulated. This level of QTL detection is an estimate of the chromosome-wise false positive rate. At a significance level of 0.05 and a map distance of 10 cM the expected number of simulation with false positive QTL was exceeded for all populations. Unlike in QTL-mapping using regression, it is not obvious how permutation tests can be implemented to determine this threshold p-value. Permutation within the smallest subclass as defined by the fixed effects might be an option. However, variance component estimation is rather time consuming given the current software. Choosing a significance level of 0.01 might be a simple solution to take multiple testing and an acceptable level of false positives into account. Alternatively, a quick method for computing approximate thresholds by Piepho [22] could be applied.

In this study the mode of inheritance was assumed to be known when a hypothesis regarding the presence or absence of a QTL were tested which is not the case in real life situations. A decision tree involving sequential hypothesis testing, as described by Thompsen *et al.* [26] might be applied. However, it seems more logical to take the 'full model' (comparable to our model 3) as the *a priori* model for gene expression. When the paternal component is not significantly different from the maternal component, the hypothesis of a Mendelian inheritance could be accepted. When they are different the sequential hypothesis testing of the full model against a model with either the maternal or paternal component only, could be applied.

4.3. Additional value of LD information

Variance methods which fit QTL as random effects can account for complex relationships between individuals in outbred populations [8, 11]. LD takes historic recombinations into account while linkage analysis adds information especially for regions with a low marker density (>10 cM).

The size of a region that is IBD can be calculated from $c = 1/(2 * \text{mutation age})$ [13]. In this study c is $1/(2 * 100) = 0.005$ M, *i.e.* 0.5 cM. The probability that two random drawn haplotypes contain such a region is given by $P_{ibd} = 1/(4 * N_e * c + 1)$, *i.e.* in this study P_{ibd} equals about five percent. Additional value of LD is therefore limited for the 10 cM map as was observed in this study. The result based on the 2 cm map showed that LD information is of added value confirming the results obtained by Lee and van der Werf [13]. The map distance used in that study was 1 cM between markers. However, up to now genome scans with multi allelic markers mapped at less than 10 cM in pigs are rare.

The advantage of using LD information can also be observed for populations with 48 sires and a 2 cM map. This population shows the largest difference between LDLA and LA-only results both in power and accuracy of location. The effect of the LD information is that the effective number of haplotypes is reduced due to the covariance among haplotype when applying LDLA. In LA-only analysis the same number of families is segregating but few offspring are available to estimate the haplotype effects.

4.4. Variance components

Using LDLA resulted in most cases in an overestimation of the QTL variance component as is shown in Table V. This was not the case for most of the LA-only analysis. The difference between LDLA and LA-only is that the covariance among the parental haplotypes is assumed to be zero for LA-only while marker and pedigree information is used to estimate this covariance in the case of LDLA. Including this covariance seems to have the effect that part of the polygenic variance is accounted for by the QTL component. This seems to be the case especially for P_{mend} and P_{pat} (Tab. V). The shifting of polygenic variance to QTL variance is more pronounced for simulations where a significant QTL was detected. Simulation with a more significant QTL showed an even more overestimated QTL component while the polygenic component was more underestimated.

In retrospect the overestimation of the QTL component(s) when LDLA was applied could have been expected. Including genetic relationships among parental haplotypes, *i.e.* the LD part in **G**, forces the variance component method to estimate the QTL-component in a base (unrelated) generation. However, the QTL-variance in this study was set at 0.1 in the offspring generation without taking LD into account.

In this study the model with a single QTL-effect (model 2) was compared to a model which allowed separate components for paternal and maternal QTL effects. Especially if QTL are differentially expressed the latter model is more appropriate. However, differences in estimated maternal and paternal QTL components also depended to a small extent on the structure of the mapping population, *i.e.* the information used for estimating the maternal and paternal component differs. A mapping population with an equal number of maternal and paternal haplotypes would therefore be optimal. In experiments involving pigs this could be approximated by including more litters per dam.

5. CONCLUSION

The present study shows how the power of finding a QTL and locating it in a certain chromosomal region depends on the population structure and the mode of inheritance of a QTL.

Estimation of parentally imprinted QTL is more efficient in designs with large family sizes. Given a fixed number of animals that can be phenotyped, for example due to the high costs of collecting data, the number of families should be balanced with the family size. Too small number of sires (<8) should be avoided. In case a large number of families is used, *i.e.* many parents, the number of haplotypes increases which reduces the accuracy of estimating the QTL effect and thereby reduces the power to show a significant QTL and to correctly position the QTL.

It is argued that including more litters per dam sired by different boars is advantageous assuming that the phenotype of the offspring does not depend on the parity of the dam.

Including LD information is advantageous because it increases both the power to detect a QTL and the ability to position the QTL while slightly decreasing the number of false positives. This is especially true for denser maps which generate more LD information. Use of information across families reduces the importance of family size.

In most studies marker density will vary from very dense to sparse. The variance component method combining linkage disequilibrium and linkage information seems to be the appropriate choice to analyze such data sets. It also adequately handles mixtures of paternal and maternal half sibs and full sib family structures which are common in pig populations.

ACKNOWLEDGEMENTS

Financial support by the Netherlands Technology Foundation (STW) was highly appreciated. Additional support was provided by the pig breeding companies Hypor and Topigs. Reviewers are thanked for comments and suggestions for improvement.

REFERENCES

[1] Beek S. van der, van Arendonk J.A.M., Criteria to optimize designs for detection and estimation of linkage between marker loci form segregating populations containing several families, Theor. Appl. Gen. 86 (1993) 269–280.

[2] Bidanel J.P., Milan D., Iannucelli N., Amigues Y., Boscher M.Y., Bourgois F., Caritez J.C., Gruand J., Le Roy P., Lagant H., Quintanilla R., Renard C., Gellin J., Ollivier L., Chevalet C., Detection of quantitative trait loci for growth and fatness in pigs, Genet. Sel. Evol. 33 (2001) 289–309.

[3] Darvasi A., Experimental strategies for genetic dissection of complex traits in animal models, Nat. Genet. 18 (1998) 19–24.

[4] De Koning D.J., Rattink A.P., Harlizius B., van Arendonk J.A.M, Brascamp E.W., Groenen M.A.M., Genome-wide scan for body composition in pigs reveals important role of imprinting, Proc. Natl. Acad. Sci. USA 97 (2000) 7947–7950.

[5] De Koning D.J., Rattink A.P., Harlizius B., Groenen M.A.M., Brascamp E.W., van Arendonk J.A.M., Detection and characterization of quantitative trait loci for growth and reproduction in pigs, Livest. Prod. Sci. 72 (2001) 185–198.

[6] De Koning D.J., Harlizius B., Rattink A.P., Groenen M.A.M., Brascamp E.W., van Arendonk J.A.M., Detection and characterization of quantitative trait loci for meat quality in pigs, J. Anim. Sci. 79 (2001) 2812–2819.

[7] Geldermann H., Mueller E., Moser G., Reiner G., Bartenschlager H., Cepica S., Stratil A., Kuryl J., Moran C., Davoli R., Brunsch C., Genome-wide linkage and QTL mapping in porcine F_2 families generated from Pietrain, Meishan and Wild Boar crosses, J. Anim. Breed. Genet. 120 (2003) 363–393.

[8] George A.W., Visscher P.M., Haley C.S., Mapping quantitative trait loci in complex pedigrees: a two-set variance component approach, Genetics 156 (2000) 2081–2092.

[9] Gilmour A.R., Cullis B.R., Welham S.J., Thompson R., ASReml reference manual 2nd edition, Release 1.0. NSW Agriculture Biometrical Bulletin 3, NSW Agriculture, Locked Bag, Orange, NSW 2800, Australia, 2002.

[10] Grindflek E., Szyda J., Liu Z., Lien S., Detection of quantitative trait loci for meat quality in a commercial slaughter pig cross, Mamm. Genome 12 (2001) 299–304.

[11] Hoeschele I., Uimari P., Grignola F.E., Zhang Q., Gage K.M., Advances in statistical methods to map quantitative trait loci in outbred populations, Genetics 147 (1997) 1445–1457.

[12] Janss L.L.G., Heuven H.C.M., LDLA, a package to compute IBD matrices for QTL fine mapping by variance component methods, Abstracts of 56th annual meeting of EAAP (2005) p. 125.

[13] Lee S.H., Werf J.H.J. van der, The efficiency of designs for fine-mapping of quantitative trait loci using combined linkage disequilibrium and linkage, Genet. Sel. Evol. 36 (2004) 145–161.

[14] Lee S.H., Werf J.H.J. van der, The role of pedigree information in combined linkage disequilibrium and linkage mapping of quantitative trait loci in general complex pedigree, Genetics 169 (2005) 455–466.

[15] Malek M., Dekkers J.C.M., Lee H.K., Baas T.J., Prusa K., Huff-Lonergan E., Rothschild M.F., A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. I. Growth and body composition, Mamm. Genome 12 (2001) 630–636.

[16] Malek M., Dekkers J.C.M., Lee H.K., Baas T.J., Prusa K., Huff-Lonergan E., Rothschild M.F., A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. II. Meat and muscle composition, Mamm. Genome 12 (2001) 637–645.

[17] Meuwissen T.H.E., Goddard M.E., Fine scale mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci, Genetics 155 (2000) 421–430.

[18] Meuwissen T.H.E., Goddard M.E., Prediction of identity by descent probabilities from marker haplotypes, Genet. Sel. Evol. 33 (2001) 605–634.

[19] Meuwissen T.H.E., Karlsen A., Lien S., Olsaker O., Goddard M.E., Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping, Genetics 161 (2002) 373–379.

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

[21] Ovilo C., Clop A., Noguera J.L., Olivier M.A., Barragan C., Rodriguez C., Silio L., Toro M.A., Coll A., Folch J.M., Sanchez A., Babot D., Varona L., Perez-Enciso M., Quantitative trait locus mapping for meat quality in an Iberian × Landrace F2 pig population, J. Anim. Sci. 80 (2002) 2801–2808.

[22] Piepho H.P., A quick method for computing approximate thresholds for Quantitative Trait Loci detection, Genetics 157 (2001) 425–432.

[23] Rohrer G.A., Keele J.W., Identification of quantitative trait loci affecting carcass composition in swine. I. Fat deposition traits, J. Anim. Sci. 76 (1998) 2247–2254.

[24] Rohrer G.A., Keele J.W., Identification of quantitative trait loci affecting carcass composition in swine. II. Muscle and wholesale product yield traits, J. Anim. Sci. 76 (1998) 2255–2262.

[25] Shuen Lo H., Zhining Wang, Ying Hu, Yang H.H., Gere S., Buetow K.H., Lee M.P., Allelic variation in gene expression is common in the human genome, Genome Res. 13 (2003) 1855–1862.

[26] Thompsen H., Lee H.K., Rothschild M.F., Malek M., Dekkers J.C.M., Characterization of quantitative trait loci for growth and meat quality in a cross between commercial breeds of swine, J. Anim. Sci. 82 (2004) 2213–2228.

[27] Van Laere A., Ngyyen M., Braunschweig M., Nezer C., Collete C., Moreau L., Archibald A., Haley C., Buys N., Tally M., Andersson G., Georges M., Andersson L., A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig, Nature 425 (2003) 832–836.