Linkage Disequilibrium Based Methods To Map QTL In Pig Familial Populations, A Simulation Study

M. del P. Schneider[*] and H. Gilbert[*]

Introduction

Linkage Disequilibrium (LD) methods have become a promising and useful tool for fine mapping quantitative trait loci (QTLs) in outbred populations. Most available strategies assume unrelated individuals in the population, whereas individuals in livestock populations are highly connected. Methods known as LDLA were proposed to deal with relatedness and avoid spurious linkage due to hidden or ignored familial structures. The objective of the study was to compare mapping accuracy and power to detect QTLs of different LD methods in simulated pig populations with connected small families and low pedigree information.

Material and methods

Simulation. Simulated populations were obtained with the Linkage Disequilibrium with Several Options (LDSO) software by Ytournel (2008). An initial population of 100 individuals (generation 0) evolved with random mating over 100 generations. No initial disequilibrium and no selection were assumed. A 10 cM genetic map with 41 biallelic markers (SNPs) equally spaced at 0.25cM was simulated. A QTL with 5 alleles was located at position 3.85 cM, *i.e.* between markers 16 and 17. Only one QTL allele had an effect on the trait. QTL and marker alleles had equal initial frequencies and were distributed at random in the founder population. After generation 100, a two-generation population with known pedigree was created. The trait phenotypes were simulated from generation 0 as the sum of the polygenic, additive QTL, and residual effects. Genotypes were available for generation 101 (parents) and 102 (progeny). Individuals were sampled in generations 101 and 102 to built populations comprising 50 or 30 half-sib families of 3 progeny. In each population scenarios assuming an initial heritability of the quantitative trait of 0.5 or 0.25, an initial QTL effect of 1.0 or 0.5 genetic s.d., and a proportion of 50 or 20% of sires heterozygote for the QTL in generation 102 were tested. A total of 1000 simulations were carried out for each scenario. Replicates were discarded when fixation occurred for the first and the last markers, and for more than 25% of the markers. The average LD *D'* value was 0.63 two cM from the true QTL location (in 100 replicates).

Statistical analyses. Three single marker (SM) methods and two haplotype based methods were used. The SM analyses were: REG, a linear regression analysis, where the phenotypic value is regressed on the genotype; QXP, a mixed model including the SNP genotype and an infinitesimal effect, carried out with QxPak software using mixed models theory and

[*] INRA, UMR1313 GABI, 78352 Jouy-en-Josas Cedex, France

maximum likelihood estimation (Perez-Enciso and Misztal 2004); and QFAM, a family-based association test for quantitative traits derived from QTDT (Abecasis, G.R., Cardon, L.R., and Cookson, W.O. 2000) that performs a linear regression of phenotype on allele dosage and then uses permutation to account for the dependence between related individuals. This last method has been developed for human populations, *i.e.* trios and full-sibs families, and implemented in the software PLINK (Purcell, S., Neale, B., Todd-Brown, K. *et al.* 2007).

The first haplotype based method, HAPIM, models the frequency of haplotypes comprising one trait locus and two flanking makers. It is based on maximum likelihood calculation and it was designed for unrelated individuals with no family information (Boitard, S., Abdallah, J., de Rochambeau, H. *et al.* 2006). Analyses were done with the R-package HAPim. The second haplotype based method, LDLA**,** uses the approach of Meuwissen and Goddard (2000). It is a variance component method which integrates the LD information between base haplotypes in the construction of the relationship matrix among QTL allelic effects. Four marker haplotypes were considered and likelihoods were maximized with an AI-REML approach (Druet, T., Fritz, S., Boussaha, M. *et al.* 2008).

For QXP, HAPIM and LDLA the location of the marker with the maximum likelihood ratio test (full *vs*. reduced model) was retained as the most likely estimate of the QTL position. REG provides F statistics (full *vs*. reduced model) and corresponding P-values, and QFAM empirical P-values. Thus, the markers with the lowest P-values were taken as the estimate of the QTL location. Except with REG and HAPIM, some pedigree information was used to compute either the test statistic or the p-values, but only generations 101 and 102 were used.

Mean square errors (MSEs) were calculated to evaluate the accuracy of the QTL position estimation for each method in each scenario. Differences in MSEs between methods were tested with GLM (SAS) fitting method as a fixed effect and adjusting for multiple comparisons (Bonferroni). To compare powers, we first used the p_values (corrected for the number of tests) provided by the softwares. Due to high number of false positive retained with this strategy, empirical thresholds were computed as the empirical 0.95 quantile of 1000 simulations under the null hypothesis (no QTL effect).

Results and discussion

Table 1 shows the MSEs for the scenarios assuming an initial heritability of 0.5, for the two population sizes, the two QTL effects and the two proportions of heterozygote sires. Within method, MSEs increased with decreasing population size, decreasing percentage of heterozygote sires and smaller QTL effect. Mapping resolution decreased with decreasing population size. It reached 7.37 cM^2 for 10 sire families with 50% of informative sires (not showed). The mapping resolution was reduced about 2.5 and 2 times for the scenarios with a smaller QTL effect (0.5 s.d.) compared to the QTL with larger effect for the populations of 50 and 30 sires, respectively. MSEs increased when the minimum percentage of informative sires was reduced, although the differences were not that large. Larger number of families potentially less informative seemed more favorable than smaller number of more informative families under the simulated scenarios.

All methods had similar mapping accuracies for positioning QTL at its true location. MSEs between SM based methods were not statistically different, whereas HAPIM and LDLA tended to have larger MSEs. QFAM and REG perform a linear regression, so it is not surprising that they had similar results in terms of position estimations. The addition of an infinitesimal effect in the QXP model did not improve the mapping accuracy, may be due to little information contribution from small pedigrees and family sizes.

Table 1: Mean Square Errors in cM^2 of QTL location estimates for different mapping LD methods assuming an initial heritability of 0.5 for the quantitative trait[*]

QTL effect (s.d.)	1.0				0.5			
Number of sires	50		30		50		30	
Het. sires (%)	50	20	50	20	50	20	50	20
REG	2.1^{a}	2.6^{ab}	3.4^{b}	4.6^{a}	5.6^{a}	6.7^{a}	7.4^{a}	7.5^{a}
QXP	2.1^{a}	2.6^{ab}	3.4^{b}	4.5^{a}	5.6^{a}	6.7^{a}	7.3^{a}	7.5^{a}
QFAM	2.2^{a}	2.6^{ab}	3.3^{b}	4.8^{a}	5.4^{a}	6.9^{a}	7.1^{a}	7.9^{a}
HAPIM	2.0^{a}	2.5^{b}	3.7^{ab}	5.1^{a}	5.9^{a}	7.0^{a}	7.0^{a}	8.2^{a}
LDLA	2.7^{a}	3.3^{a}	4.5^{a}	5.2^{a}	6.2^{a}	6.8^{a}	7.5^{a}	7.7^{a}

[*] Bonferroni t Test. MSEs with the same letter are not significantly different

HAPIM had similar or slightly greater MSEs than the SM methods. A slightly better mapping accuracy than other methods was expected, and not observed, since contrary to other methods which test for the QTL segregation at marker positions, HAPIM estimates the position of the QTL at the center of the marker bracket, where the QTL is actually simulated. LDLA had the highest MSEs, although not always statistically different from the other methods. Boitard, S., Abdallah, J., de Rochambeau, H. *et al.* (2006) and Grapes, L., Dekkers, J.C.M., Rothschild, M.F. *et al.* (2004) had shown that LDLA is better at estimating the QTL position than SM based method, although the gain in accuracy is not that large. This unexpected poor performance must be related to the convergence problems of the AI-REML at some of the tested positions (*e.g.* of all the positions tested, 83% and 66% reached convergence for the scenario with 50 sires and 30 sires respectively, with 50% informative sires and QTL effect of 1.0). It is known that location parameters (*e.g.* SNP modeled as fixed effect) are easier to estimate than dispersion parameters (*e.g.* haplotypes modeled as a random effect in LDLA), and it may have a significant effect on the estimation of the QTL position. Therefore, these results must be interpreted with caution. New simulations are currently carried out using REML to improve convergence at the expense of higher computing time.

Power to map QTL (Table 2) decreased with decreasing population size and proportion of informative sires, consistently with the MSE results. The decrease in power was considerable (about 2.5 times) for the scenarios with a QTL effect of 0.5 s.d. compared to the larger effect QTL. SM based methods had all similar power. Although LDLA had the highest MSEs, it ranked slightly better than the other methods for power, whereas HAPIM had generally lower power. Analyses were also carried out for the scenarios with 10 sires, where power

could reach 17% with the larger QTL effect. Percentage of detected QTL with the SM methods under the null hypothesis, computed from p-value, showed that the proportion of false positives was about 30-35%, with no improvement with the QFAM method compared to the REG, despite the use of permutations to account for familial structure.

Table 2: Power (in %) to detect QTL using empirical thresholds, along with percentage of false positive when using p-values, assuming initial heritability of 0.5

QTL effect (s.d.)	1.0				0.5				false positives*			
Number of sires	50		30		50		30		50		30	
Het. sires (%)	50	20	50	20	50	20	50	20	50	20	50	20
REG	86	77	67	50	33	30	23	17	35	32	31	29
QXP	85	75	66	49	32	29	22	17	36	32	34	32
QFAM	86	77	63	51	34	30	21	19	35	32	35	30
HAPIM	78	64	51	45	26	21	14	16	-	-	-	-
LDLA	87	74	63	51	39	29	24	20	-	-	-	-

* Simulations under the null hypothesis (*i.e.* no QTL), computations based on p-values

When an initial heritability of 0.25 was assumed for the quantitative trait, similar trends for MSEs and power were found (results not shown). However, the mapping accuracy and power were considerably low, *e.g.* average MSEs for the population of 50 sires, with 50% of heterozygote sires and a large QTL was 6.1 cM^2.

Conclusion

Methods performed better with increasing number of families, in more informative families and when the QTL had a large effect. Similar results were obtained with SM based regression methods and haplotype based methods, with slightly better power for LDLA. However, accounting for family relatedness did not significantly improve the power to detect the QTL nor the mapping accuracy. Further scenarios are being studied, with denser maps and deeper pedigree.

References

Abecasis, G.R., Cardon, L.R., and Cookson, W. (2000). *Am. J. Hum. Genet.*, 66:279-292.

Boitard, S., Abdallah, J., de Rochambeau, H. *et al.* (2006). *BMC Genomics,* 7:54.

Druet, T., Fritz, S., Boussaha, M. *et al.* (2008). *Genetics,* 178:2227-2235.

Grapes, L., Dekkers, J.C.M., Rothschild, M.F. *et al.* (2004). *Genetics*, 166:1561-1570.

Meuwissen, T.H. and Goddard, M.E. (2000). *Genetics*, 155:421-430.

Perez-Enciso, M., and Misztal, I. (2004). *Bioinformatics*, 20:2792-2798.

Purcell, S., Neale, B., Todd-Brown, K. *et al.* (2007). *Am. J. Hum. Genet.*, 81:559-575.

Ytournel, F. (2008*). PhD Dissertation,* Pages 219.