# Strategies for haplotype-based association mapping in a complex pedigreed population

J. Boleckova[1,2], O. F. Christensen[1], P. Sørensen[1], G. Sahana[1]

[1]Department of Molecular Biology and Genetics, Faculty of Agricultural Sciences, Aarhus University, Tjele, Denmark
[2]Department of Cattle Breeding, Institute of Animal Science, Prague-Uhříněves, Czech Republic

**ABSTRACT**: In association mapping, haplotype-based methods are generally regarded to provide higher power and increased precision than methods based on single markers. For haplotype-based association mapping most studies use a fixed haplotype effect in the model. However, an increase in haplotype length raises the number of parameters in the model, resulting in low accuracy of the estimates especially for the low-frequency haplotypes. Modeling of haplotype effects can be improved if they are assumed to be random effects, as only one parameter, i.e. haplotype variance, needs to be estimated compared to estimating the effects of all different haplotypes in a fixed haplotype model. Using simulated data, we investigated statistical models where haplotypes were fitted either as a fixed or random effect and we compared them for the power, precision, and type I error. We investigated five haplotype lengths of 2, 4, 6, 10 and 20. The simulated data resembled the Danish Holstein cattle pedigree representing a complex relationship structure and QTL effects of different sizes were simulated. We observed that the random haplotype models had high power and very low type I error rates (after the Bonferroni correction), while the fixed haplotype models had lower power and excessively high type I errors. Haplotype length of 4 to 6 gave the best results for random model in the present study. Though the present study was conducted on data structure more frequent in livestock, our findings on random vs. fixed haplotype effects in association mapping models are applicable to data from other species with a similar pedigree structure.

**Keywords**: association mapping; haplotype; complex pedigree; false positives

Several studies have shown that the analysis using marker haplotypes provides higher power and precision in quantitative trait loci (QTL) mapping than that using single markers (Akey et al., 2001, de Bakker et al., 2005). In the single marker analysis, even when the tested marker locus is in strong linkage disequilibrium (LD) with QTL, power can be quite low if the frequencies of the marker and QTL alleles are different (Kaplan and Morris, 2001). This problem is resolved using haplotype-based association methods since they fully exploit LD information from multiple markers. Both simulations (Akey et al., 2001; Zaykin et al., 2002) and empirical studies support this statement (Liu et al., 2008). Conflicting reports, however, are available (Zhao et al., 2007). The discrepancies in the results of these studies could be due to several factors influencing the LD between markers and the LD between markers and QTLs: marker density, effective population size, generations of random mating, QTL allele frequency, QTL position relative to observed marker positions, and if one of the

Supported by the Ministry of Agriculture of the Czech Republic (Projects Nos. QH91270 and QH71275) and by the Danish Agency for Science, Technology and Innovation (Grant FTP No. 09-065751).

markers is a causative mutation that affects gene products (Cargill et al., 1999; Zhao et al., 2007; Pei et al., 2009). Furthermore, imprecise modeling of the genetic relation within the study sample may cause spurious associations, i.e. when using a sire model a rare haplotype may be present in only one or a few families (Kent et al., 2007). Finally, an accurate constructing of haplotypes can also be difficult and this may reduce the power of a haplotype approach in real data (Barzuza et al., 2005; Andrés et al., 2007).

Most of the model comparisons in genome-wide association studies that used simulated or real data investigated the power and precision of QTL findings (e.g. Akey et al., 2001; Grapes et al., 2004; Grapes et al., 2006). These studies considered haplotype a fixed effect in the model. In such a model, increase in haplotype length leads to an increase in the number of effects that must be estimated. Because the frequency of some haplotypes can be very low, this could result in low accuracy of the estimates and would eliminate the benefit achieved from the improved modeling obtainable with haplotypes (Becker and Herold, 2009). Calus et al. (2008) studied the effects of haplotype length and observed that a model with 10-marker haplotypes and identity-by-descent (IBD) relationships between them yielded ~ 25 to 500 times as many effects that need to be estimated compared to a model based on a single marker. Several studies (e.g. Zhao et al., 2007; Becker and Herold, 2009; Calus et al., 2009) compared power of haplotype-based models for QTL mapping, but did not compare type I errors. Sahana et al. (2010) observed a very high type I error rate using haplotype as a fixed effect in the model. We hypothesize that loss of power and increase in type I error may be addressed by fitting haplotype as a random effect in the model. Using a random haplotype effect in association mapping has not been investigated so far in the literature. In this study, we had the two objectives: (1) to compare fixed and random haplotype models for association mapping in terms of power and type I error, and (2) to find the optimum haplotype length to maximize power and at the same time not to inflate type I error. We used a simulated dataset which resembles a population with complex pedigree relationship with a limited (historical) effective population size, in which most LD is generated by a random drift. Such populations are common for domestic animals but also include some isolated human populations (e.g. Icelandic population) and plants.

## MATERIAL AND METHODS

### Simulation of data

**Pedigree.** The data used in the study were simulated by Sahana et al. (2010) by combining a historical pedigree with the real Danish Holstein cattle pedigree. The historical pedigree was simulated from a founder population of 150 animals (75 males and 75 females), assumed to be 50 generations back in time. For each of the subsequent 50 generations, 75 males and 75 females were produced by randomly sampled parents with replacement from the previous generation and without selection. The real Danish Holstein pedigree (real pedigree) included 8500 progeny-tested bulls and was aligned with the historical pedigree as follows. The latest generation of the real pedigree was considered as generation 50 + 1. The unknown parent(s) of the animals in the real pedigree were randomly drawn from the animals of the historical pedigree generation. From the real pedigree, 2000 bulls and their sires were sampled at random with the restriction that at least 10 sons were sampled from each half-sib family. The final dataset consisted of 2069 bulls from 212 half-sib families, varying in size from 10 to 33 sons per family. Genotypes and phenotypes for these bulls were saved for the final analysis.

**Marker and quantitative trait locus alleles**. Marker alleles were sampled for 5000 biallelic loci distributed on five chromosomes (1000 markers on each chromosome) with 0.1 cM between each locus. The two alleles of each marker locus in the founder animals were sampled with equal probabilities. A total of 15 QTL were simulated on four of the chromosomes and one chromosome had no QTL. All QTL alleles in the founder animals were unique. Marker and QTL alleles were transmitted from parents to offspring for 50 generations. Recombinations were sampled according to Haldane's mapping function. Linkage disequilibrium was created by a random genetic drift. Haplotypes were known without error and no genotypes were missing.

For each QTL, one allele with a frequency between 0.10 and 0.20 was sampled at random from the QTL alleles still present in the $50^{th}$ generation. This was treated as the mutant allele, while all other alleles were combined to constitute the wild type allele. The allele substitution effects of the 15 QTL were standardized based on their allele frequencies in the last generation of the historical

pedigree, so that each QTL explained a pre-defined percentage of the genetic variance. We simulated one "big-QTL" explaining 10% of genetic variance, four "medium-QTL" explaining 5%, and ten "small-QTL" explaining 2% of it. The intervals between QTL pairs located on the same chromosome ranged from 1 to 45 cM. The QTL locations and their effect on the phenotype are presented by Sahana et al. (2010). No QTL were simulated on chromosome 5 (null chromosome).

**Phenotypes.** The phenotypes were obtained as the sum of the effects of the 15 QTL, a residual polygenic effect, and a random residual. First, the effects from the 15 QTL were summed. Then the QTL effect was standardized to a mean of zero and a variance of one. The QTL and polygenes constituted half of the total genetic variance. The residual polygenic effect was generated in two steps. First, polygenic values for the founder generation animals were sampled from a standard normal distribution. The residual polygenic values for the animals of the subsequent generations were derived by summing half of the sire and dam residual polygenic values and a Mendelian sampling term. The genetic value of an animal was the total sum of the QTL effects and the polygenic effect. The residual variance was sampled to achieve a heritability of 0.977 for bulls, corresponding to a heritability of 0.30 in individual records when the bulls had 100 daughters each with a phenotypic record. So the phenotypes were an accurate indicator of genetic merit. Half of the total heritability was explained by the 15 QTL and the remaining by the polygenes. Contrary to 25 datasets analysed by Sahana et al. (2010), just 24 replicate datasets were analysed in the present study because of general convergence problems of one dataset for both types of the models.

## Statistical analyses

Haplotypes were constructed using windows of 2, 4, 6, 10 and 20 neighbouring single-nucleotide polymorphisms (SNPs), and for a given window size the subsequent windows were overlapping. Each animal had two haplotypes, paternal (hp) and maternal (hm) in origin.

**Random haplotype model (RHM).** We used a linear mixed effect model with random polygenic effect and random effect of haplotypes. The model was:

$$y_j = \mu + u_j + q_{hm_j} + q_{hp_j} + e_j \quad (1)$$

where:

$y_j$ = phenotype of bull $j$

$\mu$ = population mean

$u_j$ = random individual polygenic effect

$q_{hm_j}$, $q_{hp_j}$ = random effects of the maternal and paternal haplotype of the animal $j$

$e_j$ = random residual effect

Individual polygenic effects were assumed to have covariances according to pedigree relationships, i.e., $u = \{u_j\}$ is normally distributed N $(0, s_g^2 \mathbf{A})$, where $s_g^2$ is the polygenic genetic variance and **A** is the additive relationship matrix derived from the pedigree. The other two random effects ($q$ and $e$) were assumed to be normally distributed with mean zero and variances $s_h\mathbf{I}$ and $s_e\mathbf{I}$, respectively, where **I** is the identity matrix. Note that the haplotype effects were the same regardless of whether they had the paternal or maternal origin, corresponding to the assumption that fathers and mothers originated from the same population. The significance of the haplotype substitution effect was assessed with a likelihood ratio test comparing the RHM model with a null-model containing mean, polygenic effect and random error terms but no haplotype effects. The analysis was performed using the DMU software package (Madsen and Jensen, 2008).

**Fixed haplotype model (FHM).** The data were reanalysed using the same linear mixed model described as in equation (1) but with the difference that haplotype was modeled as a fixed effect. The haplotype effect was assumed to be the same regardless of haplotype paternal or maternal origin. The model was specified as follows:

$$q_{hm_j} + q_{hp_j} = x_j q$$

where:

$\boldsymbol{x}_j$ = vector containing the counts of the specific haplotype [taking values 0 = not present, 1 = present in one copy (heterozygote) and 2 = homozygote for this haplotype]

$q$ = vector of haplotype effects

Note that this is different from the model with maternal and paternal haplotype effects. The significance of the haplotype effect was assessed by the Wald test using the DMU software package (Madsen and Jensen, 2008).

**Significance.** Haplotype effects were declared significant at a $P$-value of 0.05 after the Bonferroni correction, i.e. the significance threshold was chosen according to a $P$-value found by dividing 0.05

by the total number of tests carried out on the four chromosomes with simulated QTL. The midpoint of the haplotype was taken as the putative QTL position.

## Comparison of the methods

**Power.** For model comparison, QTL was declared to be correctly identified if a significant haplotype with putative QTL position within ± 2 cM of the true QTL position was found. The power is the proportion of QTL correctly identified per dataset, averaged across replicates.

**False positives.** The fifth chromosome had no simulated QTL. Therefore, any significant haplotype on this null-chromosome represented a false positive. On this chromosome, haplotype effects were declared significant at a *P*-value of 0.05 after the Bonferroni correction, i.e. the significance threshold chosen according to a *P*-value found by dividing 0.05 by the total number of tests carried out on this chromosome in one dataset. For example, in case of 10-marker haplotype, a total of 991 haplotypes were tested on the null-chomosome for one dataset, so the multiple testing correct *P*-value at 5% level was 0.05/991 = 5.045e-05. The methods were compared for type I error rate based on false positive results on the null-chromosome for each of the 24 datasets.

**Mapping precision.** The precision of a detected position was quantified as the absolute deviation of the putative QTL position compared to the true position of the simulated QTL. However, if more than one significant haplotype with a putative QTL position was found less than 2 cM from the true QTL position, precision was computed using the position of the most significant haplotype. This was done because in a real life situation the most significant marker/haplotype would be the starting point for a candidate gene search. Mean absolute errors between the true and putative positions were used to compare the models. These were also computed separately for the three categories of QTL effects. In addition, the models were also compared for percentage of identified QTL with mean absolute errors of position within the intervals of 0.5, 1.0 and 2.0 cM, separately for the three categories of QTL effects.

## RESULTS

For simplicity of presentation we abbreviated the random and fixed haplotype models as RHM and FHM respectively, and with haplotype length as suffix. For example, RHM2 means it is the random haplotype effect model with haplotype length of two markers.

**Average numbers of haplotypes.** The average numbers of haplotypes at a marker position for different haplotype lengths are presented in Table 1. The number of haplotypes at a particular position increases steeply with the increase in haplotype length. The average number of haplotypes increases from 3.97 for 2-marker haplotype to 128.86 for 20-marker haplotype. That means, while only 4 haplotype effects needed to be estimated for 2-marker haplotype in a fixed haplotype model, there were on average 129 effects needed to be estimated for 20-marker haplotype. In addition, the proportion of rare haplotypes is also very high for larger haplotypes. For example, about 95% and 85% of the haplotypes have frequency lower than

Table 1. Total number of haplotypes, their ranges and proportion of rare haplotypes at a marker position for various haplotype lengths

| Haplotype length | Total number of haplotypes at a marker position | | | Proportion of haplotypes with frequency < 0.05 | | | Proportion of haplotypes with frequency < 0.01 | | |
|---|---|---|---|---|---|---|---|---|---|
| | mean | min. | max. | mean | min. | max. | mean | min. | max. |
| 2 | 3.97 | 2 | 4 | 0.08 | 0.00 | 0.75 | 0.02 | 0.00 | 0.50 |
| 4 | 12.56 | 5 | 16 | 0.50 | 0.00 | 0.86 | 0.27 | 0.00 | 0.73 |
| 6 | 26.03 | 10 | 61 | 0.73 | 0.10 | 0.96 | 0.51 | 0.00 | 0.79 |
| 10 | 55.65 | 25 | 236 | 0.88 | 0.69 | 1.00 | 0.72 | 0.42 | 0.89 |
| 20 | 128.86 | 73 | 567 | 0.95 | 0.89 | 1.00 | 0.85 | 0.75 | 0.98 |

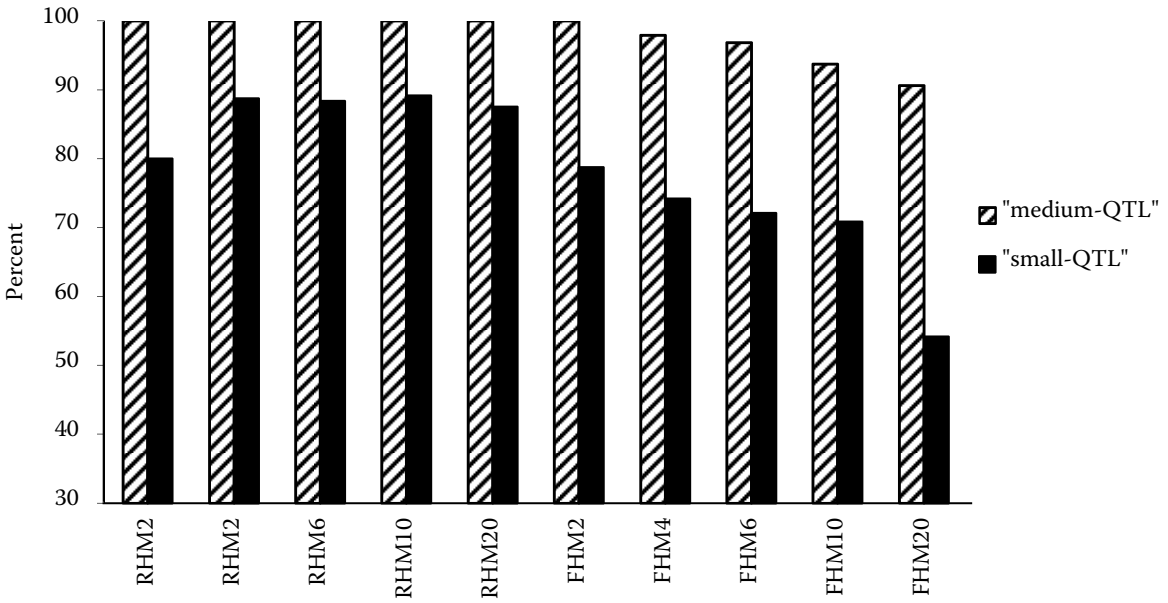


Figure 1. Power to identify QTL that explain 2% (small) and 5% (medium) of total genetic variance for random haplotype model (RHM) and fixed haplotype model (FHM) with haplotype lengths 2, 4, 6, 10, and 20

0.05 and 0.01, respectively, for 20-marker haplotype.

**Power**. The RHM detected all the "big" or "medium" QTL and the FHM identified all the "big-QTL" for all the datasets irrespective of haplotype length. However, for the FHM the power to detect the "medium-QTL" decreased with the increase in haplotype lengths (Figure 1), decreasing from 100% for FHM2 to 91% for FHM20. The differences in power between the RHM and FHM were even more distinct for detection of "small-QTL" (Figure 1). The power for the RHM increased from 80% for RHM2 to 88–89% for RHM4, RHM6, RHM10 and RHM20. In contrast, for the FHM there was a gradual loss of power with increase in haplotype length and the loss in power was much more pronounced for larger haplotypes (FHM10 vs. FHM20). For the FHM the power to detect the "small-QTL" varied from 79% (FHM2) to 54% (FHM20).

**False positives**. The numbers of significant haplotypes on the null-chromosome (false positives) by the RHM and FHM are presented in Table 2. The RHM was able to keep type I rates below expected rates of 5% after the Bonferroni correction. Among the total of 24 datasets analysed, only one dataset for RHM2, RHM10 and RHM20, two datasets for RHM6 and four datasets for RHM4 had significant haplotypes on the null-chromosome. Looking at each dataset separately, the numbers of significant haplotypes on the null-chromosome (false positives) were very small (0 to 3) for the RHM, except for one dataset, RHM20, that had 11 false positives. To the contrary, for all haplotype lengths the FHM had significant haplotypes on the null-chromosome for most of the datasets analysed. For some datasets, the FHM had also an extremely high number of false positives, e.g. FHM6 had 111 false positives for a specific dataset.

**Precision of methods**. The precision of the detected positions quantified as the mean absolute error compared to true position, are given in Figure 2. For all haplotype lengths the RHM methods had relatively smaller absolute errors in QTL location estimates than the FHM methods, in particular for the "big-QTL". For the RHM, the precision improved with an increase in haplotype length up to 6, but beyond that haplotype length the absolute error in estimating the QTL locations increased again. A similar but less pronounced pattern was observed in the FHM. Similarly, the RHM performed better than the FHM in locating QTL within a narrow interval (0.5cM) of its true position (details not presented). The difference was most pronounced for the "big-QTL". For example, 79% of the "big-QTL" was located within 0.5 cM by RHM6 while the corresponding value for FHM6 was 54%.

Table 2. Comparison of type I error between random and fixed haplotype models based on the chromosome without true QTL

| | Random haplotype model (RHM) | | | | | Fixed haplotype model (FHM) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Haplotype length | 2 | 4 | 6 | 10 | 20 | 2 | 4 | 6 | 10 | 20 |
| Total number of haplotypes tested on null-chromosome | 23 976 | 23 928 | 23 880 | 23 784 | 23 544 | 23 976 | 23 928 | 23 880 | 23 784 | 23 544 |
| Total number of FP in the 24 datasets | 1 | 4 | 4 | 3 | 11 | 311 | 668 | 795 | 763 | 578 |
| Number of datasets with FP out of the 24 datasets | 1 | 4 | 2 | 1 | 1 | 23 | 24 | 24 | 24 | 23 |
| Range of FP in one dataset | 0–1 | 0–1 | 0–3 | 0–3 | 0–11 | 0–32 | 3–78 | 2–111 | 2–109 | 0–84 |

FP = false positives, declared using $P < 0.05$ after the Bonferroni correction for multiple tests per a dataset

## DISCUSSION

We have shown that a linear mixed model with a random haplotype effect performed better both in terms of power and false positive rate compared to a fixed haplotype model. To our knowledge, earlier studies have all considered haplotype a fixed effect in the model (e.g. Zhao et al., 2007; Pryce et al., 2010).

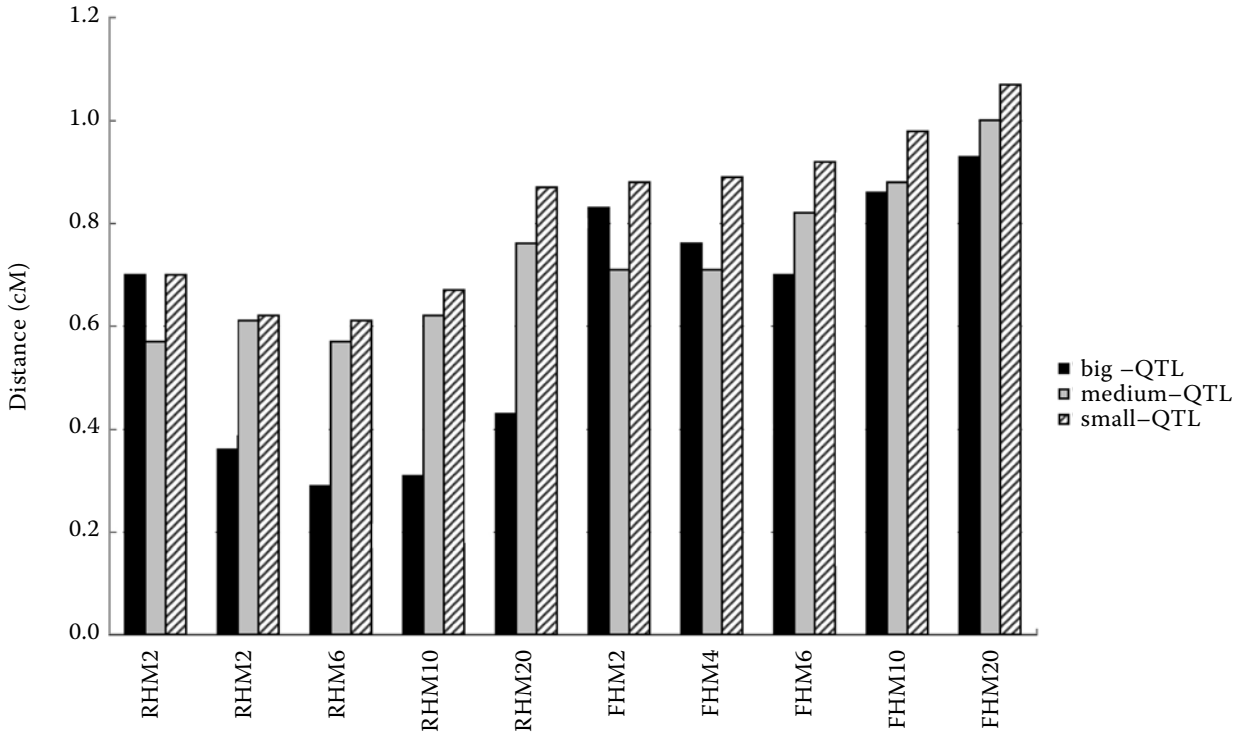


Figure 2. Average errors in estimating QTL position for the RHM and FHM models with five haplotype lengths for three categories of QTL explaining 10% (big), 5% (medium) and 2% (small) of the total genetic variance

There were two significant modeling differences in this study compared to Sahana et al. (2010) who used nearly the same datasets to compare several associations of mapping methods and models: (1) fitting haplotype effects as random in the model, and (2) modeling genetic relationship within study samples. Sahana et al. (2010) observed very high type I error when considered the half-sib family structure in the fixed haplotype based models (0.28–0.64 for various haplotype lengths). In the present study we have also observed that in spite of considering full relationship across all the animals, the fixed haplotype model resulted in Type I errors well above the nominal level.

The power of the RHM method increased up to the haplotype size of 4. This result on optimal haplotype length for QTL detection power is consistent with that of Grapes et al. (2006) and Zhao et al. (2007). The power of the RHM method appears to be amenable to improvement only up to a certain haplotype length, and further increases in haplotype length do not affect the power. This may be due to a decrease in LD between distant SNPs, which therefore add no new information. The number of parameters to be estimated in the RHM does not change with the increase in the haplotypes number. That may be the reason for not losing the power with the increase in haplotype length for the RHM, as the effect of less frequent haplotypes would be shrunk towards zero. An alternative to the RHM would be to cluster the haplotypes on the basis of their origin. The clustering could be based on IBD probabilities where haplotypes with very high IBD probability are clustered together (Blott et al., 2003; Sahana et al., 2008; Calus et al., 2009) or it could be based on genealogy (tree-based) (Pan et al., 2009). Such clustering may also be used in combination with the RHM. Further studies are needed to compare performance of the RHM with and without haplotype clustering.

Multiple SNP haplotypes in the vicinity of QTL are commonly expected to yield significant results in the association analysis. This is because sets of SNPs that are physically close to the causal factor tend to be in linkage disequilibrium. This effect declines with genetic distance and also depends on minor allele frequencies. Hence, an isolated significant SNP will often represent a spurious association or a wrongly mapped SNP. The false positives observed with the RHM were either a single significant haplotype or a few significant ones. In real studies this type of false positives will not be selected for further follow up study. However, the FHM had several significant haplotypes on the null chromosome (as many as 111), and some of them were clustered together. Such false positive results may potentially mislead an investigator as concerns the searching area of the candidate gene or candidate polymorphism.

In this study, the pedigree was known (recorded), no genotypes were missing, the haplotypes were known with certainty and, above all, the phenotypes were an accurate indicator of genetic merit. However, for real data some genotypes will be missing and haplotype construction will be inexact compared to the simulated datasets. Mistakes might also exist in the pedigree records for real data. These facts would lower the power of the proposed QTL detection method in practice, compared to what is observed in the simulated datasets. In addition, recombination rate, historical LD and density of markers may vary across the genome, and constructing haplotypes by using the nearest fixed number of SNPs is not optimal. An alternative would be optimal subsets of SNPs (Halldórsen et al. 2004). Finally, we note that the Bonferroni correction for multiple testing used in this paper is too conservative, and more sophisticated methods (Mathias et al., 2006; Huang et al., 2007) could be used instead.

## CONCLUSION

Strategies for haplotype-based association studies were compared. We observed that models with random haplotype effects performed better in comparison to models with fixed haplotype effect in terms of power, controlling type I error and precision. We may state that a haplotype length of 4 to 6 was optimal for the marker density considered in this study.

### Acknowledgement

The authors acknowledge Per Madsen, senior scientist, Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, University of Aarhus for adapting the DMU software to output the quantities needed in the Wald test for the model with fixed haplotype effect.

## REFERENCES


Akey J., Jin L., Xiong M. (2001): Haplotypes vs. single marker linkage disequilibrium tests: what do we gain? European Journal of Human Genetics, 9, 291–300.

Andrés A.M., Clark A.G., Shimmin L., Boerwinkle E., Sing C.F., Hixson J.E. (2007): Understanding the accuracy of statistical haplotype inference with sequence data of known phase. Genetic Epidemiology, 31, 659–671.

Barzuza T., Beckmann J.S., Shamir R., Pe'er I. (2005): Typing without calling the allele: a strategy for inferring SNP haplotypes. European Journal of Human Genetics, 13, 898–901.

Becker T., Herold C. (2009): Joint analysis of tightly linked SNPs in screening step of genome-wide association studies leads to increased power. European Journal of Human Genetics, 17, 1043–1049.

Blott S., Kim J.J., Moisio S., Schmidt-Küntzel A., Cornet A., Berzi P., Cambisano N., Ford C., Grisart B., Johnson D., Karim L., Simon P., Snell R., Spelman R., Wong J., Vilkki J., Georges M., Farnir F., Coppieters W. (2003): Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics, 163, 253–266.

Calus M.P., Meuwissen T.H., de Roos A.P., Veerkamp R.F. (2008): Accuracy of genomic selection using different methods to define haplotypes. Genetics, 178, 553–561.

Calus M.P., Meuwissen T.H., Windig J.J., Knol E.F., Schrooten C., Vereijken A.L., Veerkamp R.F. (2009): Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. Genetics Selection Evolution, 41, 11.

Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N., Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L., Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q., Lander E.S. (1999): Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nature Genetics, 22, 231–238.

de Bakker P.I., Yelensky R., Pe'er I., Gabriel S.B., Daly M.J., Altshuler D. (2005): Efficiency and power in genetic association studies. Nature Genetics, 37, 1217–1223.

Grapes L., Dekkers J.C.M., Rothschild M.F., Fernando R.L. (2004): Comparing linkage disequilibrium-based methods for fine mapping quantitative trait loci. Genetics, 166, 1561–1570.

Grapes L., Firat M.Z., Dekkers J.C.M., Rothschild M.F., Fernando R.L. (2006): Optimal haplotype structure for linkage disequilibrium-based fine mapping of quantitative trait loci using identity by descent. Genetics, 172, 1955–1965.

Halldórsen B.V., Bafna V., Lippert R., Schwartz R., de la Vega F.M., Clark A.G., Istrail S. (2004): Optimal haplotype block-free selection of tagging SNPs for genomewide association studies. Genome Research, 14, 1633–1640.

Huang B.E., Amos C.I., Lin D.Y. (2007): Detecting haplotype effects in genomewide association studies. Genetic Epidemiology, 31, 803–812.

Kaplan N., Morris R. (2001): Issues concerning association studies for fine mapping a susceptibility gene for a complex disease. Genetic Epidemiology, 20, 432–457.

Kent J.W. Jr., Dyer T.D., Göring H.H., Blangero J. (2007): Type I error rates in association versus joint linkage/association tests in related individuals. Genetic Epidemiology, 31, 173–177.

Liu N., Zhang K., Zhao H. (2008): Haplotype-association analysis. Advances in Genetics, 60, 335–405.

Madsen P., Jensen J. (2008): A user's guide to DMU. A package for analysing multivariate mixed models. Version 6, release 4.7. University of Aarhus, Faculty of Agricultural Sciences (DJF), Department of Genetics and Biotechnology, Research Centre Foulum, Tjele, Denmark. Available from http://dmu.agrsci.dk/

Mathias R.A., Gao P., Goldstein J.L., Wilson A.F., Pugh E.W., Furbert-Harris P., Dunston G.M., Malveaux F.J., Togias A., Barnes K.C., Beaty T.H., Huang S.K. (2006): A graphical assessment of p-values from sliding window haplotype tests of association to identify asthma susceptibility loci on chromosome 11q. BMC Genetics, 7, 38, doi: 10. 1186/1471-2156-7-38.

Pan F., McMillan L., Pardo-Manuel De Villena F., Threadgill D., Wang W. (2009): TreeQA: quantitative genome wide association mapping using local perfect phylogeny trees. Pacific Symposium on Biocomputing, 14, 415–426.

Pei Y.F., Zhang L., Liu J.F., Deng H.W. (2009): Multivariate association test using haplotype trend regression. Annals of Human Genetics, 73, 456–464.

Pryce J.E., Bolormaa S., Chamberlain A.J., Bowman P.J., Savin K., Goddard M.E., Hayes B.J. (2010): A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. Journal of Dairy Science, 93, 3331–3345.

Sahana G., Lund M.S., Andersson-Eklund L., Hastings N., Fernandez A., Iso-Touru T., Thomsen B., Viitala S., Sø-

rensen P., Williams J.L., Vilkki J. (2008): Fine-mapping QTL for mastitis resistance on BTA9 in three Nordic red cattle breeds. Animal Genetics, 39, 354–362.

Sahana G., Guldbrandtsen B., Janss L., Lund M.S. (2010): Comparison of association mapping methods in a complex pedigreed population. Genetic Epidemiology, 34, 455–462.

Zaykin D.V., Westfall P.H., Young S.S., Karnoub M.A., Wagner M.J., Ehm M.G. (2002): Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Human Heredity, 53, 79–91.

Zhao H.H., Fernando R.L., Dekkers J.C.M. (2007): Power and precision of alternate methods for linkage disequilibrium mapping of quantitative trait loci. Genetics, 175, 1975–1986.

Received: 2011–03–11
Accepted after corrections: 2011–05–17

*Corresponding Author*

Ing. Mgr. Jana Bolečková, Ph. D., Institute of Animal Science, Department of Cattle Breeding, Přátelství 815, 104 00 Prague-Uhříněves, Czech Republic
Tel. +420 267 009 540, e-mail: boleckova.jana@seznam.cz