Journal of Animal and Veterinary Advances 8 (12): 2626-2630, 2009
ISSN: 1680-5593

Genomic Scan to Detect QTL Using SNP Markers for Simulated Data by Regression Analysis in Half-Sib Design

[1]Hasan Koyun, [2]Seyrani Koncagül and [1]Hayrettin Okut
[1]Department of Animal Science, Faculty of Agriculture,
Biometry and Genetics Unit, Yuzuncu Yil University, 65080 Van, Turkey
[2]Department of Animal Science, Faculty of Agriculture,
Animal Breeding and Genetics Unit, Harran University, 63100 Sanliurfa, Turkey

Abstract: The aim of the present study was to conduct a genome-wide screening for QTL (Quantitative Trait Loci) detection using the simulated phenotype and genotype data sets obtained from the QTL-MAS workshop 12. A genome scan was carried out in 45 half-sib families to identify QTL influencing a hypothetical trait. Among six chromosomes, each chromosome with 1000 SNP loci, 11 informative markers at least 2 cM apart from one another were chosen based on PICs with highest χ^2-statistic. Half-sib data were pooled and used simultaneously in the analyses conducted for each chromosome separately. Data were analyzed by generating an F-statistic every 1 cM on a linkage map by regression of phenotypes on the probabilities of inheriting an allele from the sire. Permutation tests at chromosome-wide significance thresholds were carried out over 1000 iterations. Among six chromosomes, significant putative QTL were detected on chromosome 1 (27 cM), 2 (36 cM), 3 (18 cM), 4 (0 cM) and 5 (96 cM) across the families ($\alpha = 0.01$ and $\alpha = 0.05$). There was no QTL detected, exceeding chromosome-wide significance level of $p<0.05$ and $p<0.01$ on chromosome 6.

Key words: Simulated SNP markers, genome scan, regression, quantitative trait loci, interval mapping

INTRODUCTION

SNPs (Single Nucleotide Polymorphisms) are DNA sequence variations at the base of a single nucleotide (A, T, C, or G) in the genome differing between members of a species. SNPs are generally bi-allelic genetic markers that tend to be less polymorphic than RFLP (Restriction Fragment Length Polimorphism), SSR (Simple Sequence Repeat) and other multi allelic genetic markers although, the most abundant class of DNA polymorphisms (Nielsen, 2000; Heaton *et al*., 2002).

As genetic markers, SNPs are essential for development of genetic test for economically important traits in livestock and complex disease-related traits in other species as well as in humans. Due to being stable through many generations, SNPs can be used for establishing ancestral relationships among individuals. Taking these relationships into account one can reconstruct haplotypes for any region of interest in the genome (Cervino *et al*., 2005; Stone *et al*., 2005).

Significant associations of SNP marker alleles with the phenotype in question suggest linkage of the marker to QTL (Quantitative Trait Loci), or the QTL with major effect on phenotype containing that particular SNP. Additionally, in the last few years, there has been increase in population-based studies to identify genomic regions that are associated with common diseases in humans and economically important quantitative traits of livestock and plants (Cleves, 2005). Moreover, a genome-wide search for QTL associations using SNPs will become more popular and cheaper for genotyping of individuals than using conventional genetic markers such as microsatellites with regards to DNA (SNP) microarrays or chips in the near future.

The aim of the study was to perform a genomic scan thus, detecting or identifying putative QTL locations along with chromosomes using the simulated data sets obtained from the QTL-MAS workshop 12.

MATERIALS AND METHODS

Data set: Simulated common data sets including phenotypic and genotypic data obtained from the QTL-MAS workshop 12. The phenotypic data set consists of 5,865 individuals from seven generations. The genotype file contains the genotype of each animal in the pedigree described in the phenotype file. There are 6,000 loci evenly distributed over 6 chromosomes (1,000 markers/chromosome), with 0.1 centi-Morgan (cM) between markers.

Corresponding Author: Hasan Koyun, Department of Animal Sciences, Faculty of Agriculture, Biometry and Genetics Unit, Yuzuncu Yil University, Zeve Kampus, 65080 Van, Turkey

Statistical methods: The common data sets were analyzed and evaluated using a web-based GridQTL (http://gridqtl.cap.ed.ac.uk:8080/gridsphere) computer program developed by Seaton *et al.* (2002) and SAS-statistical software package (version 9.13). Analysis of SNP alleles for choosing SNPs with the highest PIC (Polymorphism Information Content) values was performed using the Allele Procedure function of SAS. SNPs and trait associations within and across families were detected using GridQTL with a single QTL model.

QTL analysis: Offspring of the 43-45 sires from the simulated data spanning 4 generations were evaluated. First of all, 11-13 informative SNPs from each of the six chromosomes containing 1000 SNP loci at 0.1 cM distance in adjacent each other were chosen based on the highest χ^2-statistics values of Polymorphism Information Content (PIC values of SNP markers) by using allele procedure of SAS-statistical software package (version 9.13). Then, the method of Haley and Knott (1992), Knott *et al.* (1996) and De Koning *et al.* (1998, 2001) were adopted for the detection and mapping of QTL in half-sib families using least square simple regression analysis.

The half-sib model of GridQTL runs within and across sire families. The analysis carried out in a two-step procedure. Firstly, SNP marker data on progeny of common parent (sire) were combined in a multipoint approach to obtain the probability of inheriting an allele or the other from the sire at particular region of a chromosome of interest. The calculated probabilities were combined into coefficients with values varying between 0 and 1. Secondly, the phenotypic values on half-sib family members were regressed on these coefficients in a within-common sire regression analysis. A linear model with the fixed effects of generation and sex was fitted to coefficients and phenotypic data. Appropriate F-statistic thresholds for a $p<0.05$ and $p<0.01$ chromosome-wide type 1 error level were generated by permutation test as described by Churchill and Deorge (1994) and Deorge and Churchill (1996). The significant threshold levels and F-statistics (for $p<0.05$ and $p<0.01$) were computed by GridQTL program. When the F-statistic exceeded the F-threshold value, it was indicated as a SNP-trait association. The analyses were carried out for each chromosome separately.

RESULTS AND DISCUSSION

Table 1 shows, the selected informative SNP markers for each chromosome and the number of sire families used in the analysis. Although, >13 informative SNP markers were identified by SAS program for each chromosome, 11-13 SNP markers were chosen among them for the purpose of ease of computation by GridQTL computer program. Another reason for not choosing more SNP markers was to guaranty the distance between the adjacent SNP marker loci being at least 2 cM apart. The number of allele ranged from 1-2 for each SNP locus.

Table 2 and Fig. 1 (a-f) show that the PIC values of SNP markers vary depending upon chromosomes. The highest SNP-PIC value (0.686) was obtained on chromosome 1 whereas the lowest SNP-PIC value (0.273) was seen on chromosome 2.

Table 3 shows, the estimated QTL locations corresponding to the peak of F-statistics, as well as results of chromosome-wide analysis with 5 and 1% thresholds for the phenotype across sires. Significant QTL locations were identified between intervals for chromosome 1 (9-47 cM), 2 (18-55 cM), 3 (0-50 cM), 4 (0-54 cM) and 5 (64-100 cM) and putative QTL influencing trait were detected in 27, 36, 18, 0 and 96 cM positions on each chromosome, respectively (α = 0.01

Table 1: Selected SNP markers according to PICs based on χ^2-test using allele and haplotype procedure in SAS statistical software for each chromosome

	Chromosomes					
Selected SNPs	1	2	3	4	5	6
1	SNP1*	SNP3	SNP3	SNP2	SNP1	SNP2
2	SNP13	SNP102	SNP89	SNP104	SNP101	SNP104
3	SNP33	SNP199	SNP195	SNP209	SNP200	SNP201
4	SNP128	SNP249	SNP289	SNP326	SNP304	SNP311
5	SNP206	SNP378	SNP432	SNP405	SNP403	SNP433
6	SNP266	SNP545	SNP525	SNP507	SNP508	SNP507
7	SNP396	SNP616	SNP621	SNP643	SNP636	SNP625
8	SNP507	SNP730	SNP731	SNP716	SNP715	SNP716
9	SNP596	SNP852	SNP823	SNP848	SNP814	SNP787
10	SNP698	SNP903	SNP920	SNP921	SNP915	SNP864
11	SNP724	SNP1000	SNP992	SNP992	SNP996	SNP921
12	SNP732	-	-	-	-	SNP992
13	SNP996	-	-	-	-	-
Total SNPs	13	11	11	11	11	12
Total family	45	44	44	44	45	43

*SNPi = SNP i^{th} cM apart from the left end of the corresponding chromosome

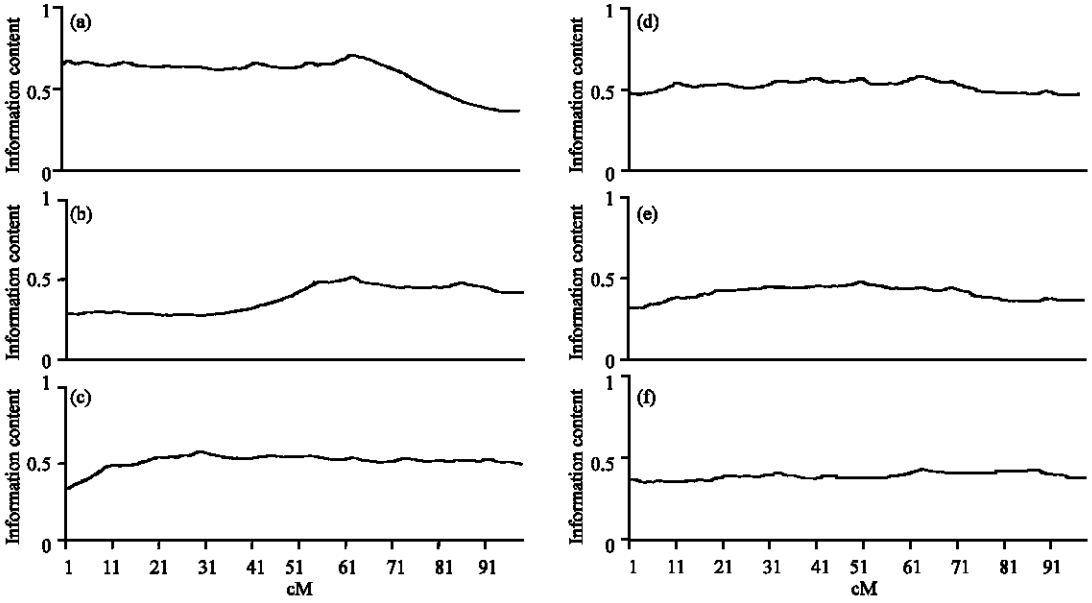

Fig. 1: Polymorphism Information Content (PIC) values and positions of SNP markers along with (a-f) chromosomes 1-6

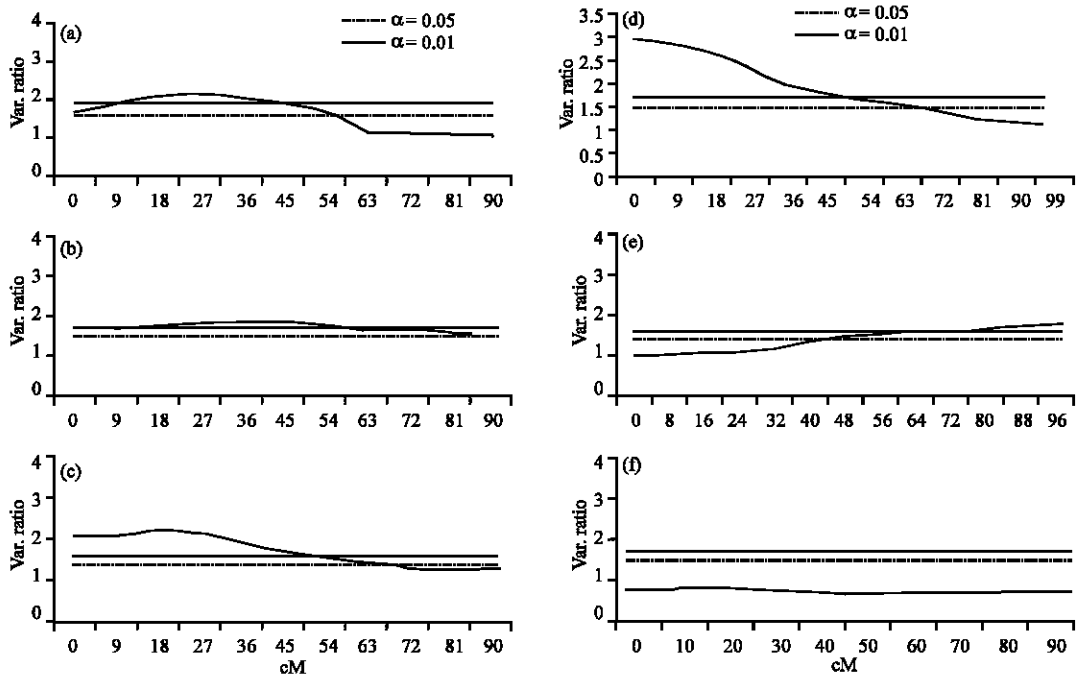

Fig. 2: Genome scan of each chromosome to identify putative QTL influencing trait positions along with (a-f) chromosomes 1-6

and α = 0.05). There were no QTL detected, exceeding chromosome-wide significance level of $p<0.05$ and $p<0.01$ on chromosome 6. Entire genomic scan and significant levels of ($p<0.05$ or $p<0.01$) chromosomal regions scanned chromosome-wide were shown in Fig. 2 (a-f).

This study presents a pioneering example of a genome-wide, QTL detection method using simulated SNP markers, as well as phenotypic and genotypic data for traits in question. Based on a chromosome-wide screening protocol, it was concluded that SNP markers detected on

Table 2: Lowest and highest PIC values of SNP markers and their positions (in cM) on chromosomes

	Chromosome					
PIC values/position	1	2	3	4	5	6
Lowest	0.324	0.273	0.343	0.463	0.312	0.345
Position (cM)	97	25	0	98	0	7
Highest	0.686	0.505	0.584	0.582	0.477	0.431
Position (cM)	61	61	29	64	51	62

Table 3: The estimated QTL locations corresponding to the peak of F-statistics and chromosome-wide 5 and 1% thresholds for the phenotype for each chromosome

	Chromosome					
Statistical analysis	1	2	3	4	5	6
QTL at (cM)	27	36	18	0	96	-
Interval (cM)	9- 47	18- 55	0- 50	0-54	64- 100	-
Critical F ($\alpha = 0.05$)	1.5685	1.5165	1.5371	1.5547	1.5480	1.5427
Critical F ($\alpha = 0.01$)	1.8594	1.7298	1.7716	1.7648	1.7542	1.7448
Highest F-value	2.16	1.91	2.5	3.08	1.99	0.87
LR*	96.41	83.69	109.03	133.94	89.03	37.49

*LR = Likelihood Ratio

chromosome 1 (27 cM), 2 (36 cM), 3 (18 cM), 4 (0 cM) and 5 (96 cM) had significant putative QTL associations. However, detection of QTL-influencing traits that was based on an F statistic computed from sums of squares explained only additive effects. Accordingly, such detection does not contain dominant and epistatic effects, it also does not explain additive, dominance and epistasis coefficients, resulting in the estimation of QTL locations along the chromosomes with a wide confidence interval.

As pointed out earlier, the results presented here are from an initial genomic search that will enable the performance of further fine-mapping analysis of putative QTL-affecting traits using multi-QTL model(s) for detection of both QTL and QTL-SNP associations per chromosome. Consequently, in order to make more precise estimations of QTL locations throughout the genome, calculations should be carried out considering >1 QTL at a time and with additive, dominant and epistatic effects and interactions determined as well.

ACKNOWLEDGEMENTS

For supplying and permitting to use the common data sets, we would like to specially express the appreciation and acknowledge the QTL-MAS 2008 workshop organizing committee; L. Ronnegard, F. Besnier, J.M., Alvarez- Castro, W. Ek, A. Johansson, L. Crooks and M. Petterson, the Group of Computational Genetics leaded by Örjan Carlborg from Department of Animal Breeding and Genetics of the Swedish University of Agricultural Sciences, Uppsala, Sweden. We would like to also thank Jules Hernandez-Sanchez from Institute of Evolutionary Biology, University of Edinburg, UK for his precious help and advice to run GridQTL program and Ben Hayes from Department of Primary Industries, Victoria, Australia for his informative assistance regarding to QTL analyses.

REFERENCES

Cervino, A.C., G. Lia, S. Edwardsa, J. Zhua, C. Lauriea, G. Tokiwaa, P.Y. Luma, S. Wangb, L.W. Castellinib, A.J. Lusisb, S. Carlsona, A.B. Sachsa and E.E. Schadt, 2005. Integrating QTL and high-density SNP analyses in mice to identify Insig2 as a susceptibility gene for plasma cholesterol levels. Genomics, 86(5): 505-517. DOI: 10.1016/j.ygeno.2005.07.010. http://www.sciencedirect.com/science?_ob=ArticleListURL&_method=list&_ArticleListID=982142706&_sort=r&view=c&_acct=C000040898&_version=1&_urlVersion=0&_userid=736695&md5=647134aa9cd427bba32e943abf45edf1.

Churchill, G.A. and R.W. Deorge, 1994. Empirical threshold values for quantitative trait mapping. Genetics, 138: 963-971. PMID: 7851788. http://www.genetics.org/cgi/reprint/138/3/963.

Cleves, M.A., 2005. Exploratory analysis of Single Nucleotide Polymorphism (SNP) for quantitative traits. The Stata J., 5 (2): 141-153. http://biostat-resources.com/stata/qtlsnp/qtlsnp.pdf.

Deorge, R.W. and G.A. Churchill, 1996. Permutation tests for multiple loci affecting a quantitative character. Genetics, 142: 285-294. http://www.genetics.org/cgi/reprint/142/1/285.

De Koning, D.J., P.M. Visscher, S.A. Knott and C.S. Haley, 1998. A strategy for QTL detection in half-sib populations. Anim. Sci., 67: 257-268. http://genepi.qimr.edu.au/contents/p/staff/CVPV032.pdf.

De Koning, D.J., N.F. Schulmant, K. Elo, S. Moisio, R. Kinos, J. Vilkki and A. Maki-Tanila, 2001. Mapping of multiple quantitative trait loci by simple regression in half-sib designs. J. Anim. Sci., 79: 616-662. http://jas.fass.org/cgi/reprint/79/3/616?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&searchid=1&FIRSTINDEX=0&volume=79&firstpage=616&resourcetype=HWCIT.

Haley, C.S. and S.A. Knott, 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Genetics, 132: 1211-1222. PMID: 16718932. http://www.genetics.org/cgi/reprint/132/4/1211.

Heaton, M.P., G.P. Harhay, G.L. Bennett, R.T. Stone, W.M. Grosse, E. Casas, J.W. Keele, T.P.L. Smith, Chitko C.G. Mckown and W.W. Laegreid, 2002. Selection and use of SNP markers for animal identificatin and paternity analysis in U.S. beef cattle. Mamm. Genome, 13: 272-281. DOI: 10.1007/s00335-001-2146-3. PMID: 12016516. http://www.springerlink.com/content/h8fqj6vdby4nwrtb/fulltext.pdf.

Knott, S.A., J.M. Elsen and C.S. Haley, 1996. Methods for multiple marker mapping on quantitative trait loci in half-sib populations. Theor. Appl. Genet., 93: 71-80. DOI: 10.1007/BF00225729.

Nielsen, R., 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics, 154: 931-942. http://www.genetics.org/cgi/reprint/154/2/931.

Seaton, G., C.S. Haley, S.A. Knott, M. Kearsey and P.M. Visscher, 2002. QTL Express: Mapping quantitative trait loci in simple and complex pedigrees. Bioinformatics, 18: 339-340. PMID: 11847090. http://bioinformatics.oxfordjournals.org/cgi/reprint/18/2/339.

Stone, R.T., E. Casas, T.P.L. Smith, J.W. Kele, G. Harhay, G.L. Bennett, M. Koohmaraie, T.L. Wheeler, S.D. Shackelford and W.M. Snelling, 2005. Identification of genetic markers for fat deposition and meat tenderness on bovine chromosome 5: Development of a low-density single nucleotide polymorphism map. J. Anim. Sci., 83: 2280-2288. http://jas.fass.org/cgi/reprint/83/10/2280.