Investigation on presence of major gene for body weight, feed intake and feed efficiency using a segregation analyses in a mice population

Burak KARACAÖREN[a,✉], Issam MOSLEM[b]

Akdeniz University, Faculty of Agriculture, Department of Animal Science, Antalya, Turkey.
[a]ORCID: 0000-0003-2981-6540; [b]ORCID: 0000-0001-8994-6916.

Abstract: Recent developments in molecular genetics and statistics have allowed the identification and use of major genes to explain the genetic variation. In this context, segregation analysis is a fast, reliable and inexpensive method that uses only phenotype and pedigree information. The aim of this study was to examine whether body weight, feed intake, and feed efficiency in a mouse population are directed by major gene in addition to polygenic and major gene effects by segregation analysis. For this purpose, previously collected dataset was used (n= 661). In this study, genetic variance, error variance, major gene variance, additive and dominant gene effects were estimated by segregation analyses. Dominant variance (1.04) was found to be smaller than the additive genetic variance (7.32) for body weight. Polygenic and major gene heritability predicted as 0.29 ($\pm$ 0.63) and 0.81 ($\pm$ 0.98) for body weight, 0.35 ($\pm$ 0.63) and 0.96 ($\pm$ 0.98) for feed intake and 0.52 ($\pm$ 0.63) and 0.81 ($\pm$ 0.98) for feed efficiency respectively. Existence of major gene was determined by examining the highest probability density regions. Although the major gene has been identified for body weight and feed intake, this result is not confirmed by the Mendelian transmission probabilities.
Keywords: Bayesian analyses, gibbs sampling, major gene, quantitative phenotype, segregation analyses.

Farelerde vücut ağırlığı, yem tüketimi ve yem etkinliğinden sorumlu major gen varlığının segregasyon analizi ile incelemesi

Özet: Moleküler genetik ve istatistikte elde edilen son gelişmeler sayesinde genetik çeşitliliğin açıklanmasında major genlerin belirlenmesi ve kullanılması olanaklı hale gelmiştir. Segregasyon analizi sadece fenotip ve pedigri bilgilerini kullanan hızlı, güvenilir ve ucuz bir yöntemdir. Bu çalışmanın amacı bir fare populasyonunda vücut ağırlığı, yem etkinliği ve yem tüketiminin poligenik etkilere ek olarak major gen tarafından yönlendirilip yönlendirilmediğinin segregasyon analizi ile incelenmesidir. Bu amaçla daha önceden toplanan bir veri seti kullanılmıştır (n= 661). Bu çalışmada segregasyon analizi ile eklemeli genetik varyans, hata varyansı, major gen varyansı, eklemeli ve baskın gen etkileri tahmin edilmiştir. Vücut ağırlığında; dominant etkinin (1.04) eklemeli gen etkisinden (7.32) daha küçük olduğu bulunmuştur. Poligenik model kullanılarak kalıtım derecesi tahmini 0.29 ($\pm$ 0.63) ve major gen modelinde kalıtım derecesi 0.81 ($\pm$ 0.98) olarak bulunmuştur. Poligenik model kullanılarak kalıtım derecesi tahmini 0.35 ($\pm$ 0.63) ve major gen modelinde kalıtım derecesi 0.96 ($\pm$ 0.98) olarak bulundu Poligenik model kullanılarak kalıtım derecesi tahmini 0.52 ($\pm$ 0.63) ve major gen modelinde kalıtım derecesi 0.81 ($\pm$ 0.98) olarak bulundu. Major gen varlığı en yüksek soncul olasılık bölgeleri incelenerek belirlenmiştir. Her ne kadar vücut ağırlığı ve yem tüketimi için major gen belirlenmiş olsada, mendelci geçiş olasılıkları bu sonucu doğrulamamıştır.
Anahtar sözcükler: Bayesçi analiz, gibbs örneklemesi, kantitatif fenotip, major gen, segregasyon analizi.

Introduction

The infinitesimal model in quantitative genetics postulate that a quantitative trait is determined by large number of loci with every locus has tiny effect (2). However genomic studies showed that in addition to polygenic effects sourced from the infinitesimal model there might be loci with also larger effects referring to the genetic architecture of the quantitative trait (5). It might be beneficial to detect such loci with larger effects (as such a major gene) for obtaining efficient selection and breeding programs in animal science. Although genotyping of an animal costs decrease however applying genomic studies at the farm level, still is not affordable at developing countries. In that regard, segregation analyses would be a cheap and accurate methodology to detect major genes (9) for animal breeding at developing countries. It would be informative to compare the results of segregation analyses with the output of genomic

analyses to evaluate if both methodologies lead to the same conclusion (15).

Recent advances in biotechnology have made it possible to investigate the genome at the molecular level. Ehsani et al. (3) and Karacaören (12) conducted genome wide association studies to infer genomic architecture of the body weight, feed intake and feed efficiency in mice using deoxyribonucleic acid (DNA), (or molecular marker) information. However, DNA information might not be available in most of the breeding populations due to the technical and/ or economical reasons especially at the developing countries. The aim of this study was to investigate existence of a major gene for body weight, feed intake and feed efficiency using only pedigree and phenotype information in a mice population in which major gene(s) have been confirmed by using DNA information in a previous study (3). Detection of major gene is important for a wide range of scientific and industrial process. Recent developments in the field of genomics have led to a renewed interest in detection of major gene(s) in DNA level. There has been no detailed investigation of comparisons for segregation and genomic analyses results. To our knowledge this is the first study to compare the output of genome wide association study with the segregation analyses.

Material and Methods

Materials: An F_2 population ($n = 661$; 323 males, 338 females) was formed by crossing M16 (F0; $n =12$) and ICR (F0; $n=12$) mouse lines for body weight studies at 8 weeks of age (1, 3). The pedigree file included 11 half-sib families. The M16 line was created by selecting for rapid weight gain while the ICR line was used as random control. More details about the data set could be found at (3).

Methods: A complex segregation model (9) incorporating both polygenic and major gene components was used to investigate existence of a major gene for the body weight, feed intake and feed efficiency in mice. Following mixed inheritance model was used to detect major gene in relation with body weight, feed intake and feed efficiency:

$$y = X\beta + Zu + ZWm + e \qquad [1]$$

where y is the vector of phenotypes (body weight, feed intake, feed efficiency), β is a vector of fixed effects of sex, u is a random vector of individual polygenic effects, W is a design matrix that includes the genotype of each mouse, m is a vector of genotypic means, e is a vector of residuals, and X and Z are incidence matrices connecting the phenotypes with model parameters. Two alleles are assumed at the major gene locus: by genotypes of AA, AB and BB with Mendelian transmission probabilities and associated additive (*a*), and dominance effect (*d*) (4).

Poligenic model and associated inference could be obtained by omission of the ZWm term from the model [1] Distributional assumptions for polygenic effects were $u|A, \sigma_u^2 \sim N(0, A\sigma_u^2)$, where A is obtained from the relationship matrix of the pedigree file with polygenic variance of σ_u^2. Residuals were assumed to be normally distributed with error variance (σ_e^2) of $e \sim N(0, \sigma_e^2)$. Uniform prior distributions were used for the model parameters of β and m in [1]. The variances are estimated using inverse chisquare prior distributions. The Gibbs sampling strategy with blocking of parents and progeny algorithm was used to obtain the desired posterior distributions by Bayesian segregation analyses using iBay software (8). Polygenic model heritability determined by $\frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$ and major gene heritability determined by $\frac{\sigma_u^2 + \sigma_a^2}{\sigma_u^2 + \sigma_e^2 + \sigma_a^2}$. where σ_a^2 is major gene variance. 10 replicates run of the Markov chain for each trait consisted of 100.000 samples, and there after each 10th sample was collected to obtain 10000 samples because of the high correlation among the samples. Gelman et al. (6) propose a convergence diagnostic for markov chain monte carlo algorithms by comparing within (W) and between (B) chain ($\psi_{ij} (i = 1, …, n;\ j=1,…,m)$) variances:

$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\bar{\psi}_{.j} - \bar{\psi}_{..})^2, \qquad \text{where}$$

$$\bar{\psi}_{.j} = \frac{1}{n} \sum_{i=1}^{n} \psi_{ij} \ , \ \bar{\psi}_{..} = \frac{1}{m} \sum_{j=1}^{m} (\bar{\psi}_{.j})$$

$$W = \frac{1}{m} \sum_{j=1}^{m} s_j^2, \ \text{where } s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\psi_{ij} - \bar{\psi}_{.j})^2.$$

with number of iterations, *n*, and replicates, *m*, leads to variances of the chain as

$$\widehat{\text{var}}^+(\psi \backslash y) = \frac{n-1}{n} W + \frac{1}{n} B \ .$$

By using $\widehat{\text{var}}^+(\psi \backslash y)$ $\hat{R}$ test statistics could be used to investigate convergence of the chain by if the limiting value of $\hat{R}$ approach to 1 as n→ ∞.

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}^+(\psi \backslash y)}{W}} \ .$$

Results

The assumption of normality was tested by Kolmogrow-Smirnow test for all the phenotypes and confirmed the normality for feed intake (P> 0.05). The results of segregation analyses of body weight are given in Table 1. The polygenic variance was found to be smaller than the major gene variance (Figure 1 and Figure 2). Probably deviation from normality confounded with major gene variance hence it could have overestimated (Table 1). We applied various logarithmic and geometric data transformations to body weight to obtain normality

(results are not shown) but still, major gene variance found to be larger. Existence of the major gene for body weight is confirmed by the 95% HPDR, not including zero (Table 1). In terms of frequentist statistical inference, rejecting the null hypothesis stating no major gene for body weight at 95 % significance level. Dominance effect was found to be smaller (1.04) than the additive effect (7.32). Results of 95 % HPDR for Mendelian transmission probabilities are given in Table 2. Since 95 % HPDR is not included

Mendelian transmission probabilities of 1, 0.5 and 0 it was concluded that mode of inheritance of body weight is not Mendelian. Estimates of major gene heritability for body weight were found to be around 0.81 while polygenic heritability found to be 0.29. Convergence analyses of the Gibbs sampler based on $\hat{R}$ given at Table 4. Convergence was concluded since estimates of $\hat{R}$ found to be around 1 for all model parameters.

Table 1. Posterior mean and standard deviation and left and right bounds of the 95% highest posterior density (HPDR 95 %) for body weight, and feed intake.

Parameter	Body weight				Feed intake			
	Posterior Mean	Posterior Standard Deviation	HPDR 95 % left	HPDR 95 % right	Posterior Mean	Posterior Standard Deviation	HPDR 95 % left	HPDR 95 % right
Error variance	11.13	1.96	6.56	17.34	142.22	37.21	129.01	272.07
Polygenic variance	4.46	2.55	0.00	18.52	132.29	58.89	111.34	423.27
Major gene variance	44.21	121.37	101.71	1401.92	919.99	2111.35	20355.53	382302.40
Additive effect	7.32	11.61	0.00	61.24	28.80	26.70	0.00	162.81
Dominant effect	1.04	6.38	0.00	37.87	24.88	20.19	41.84	125.20

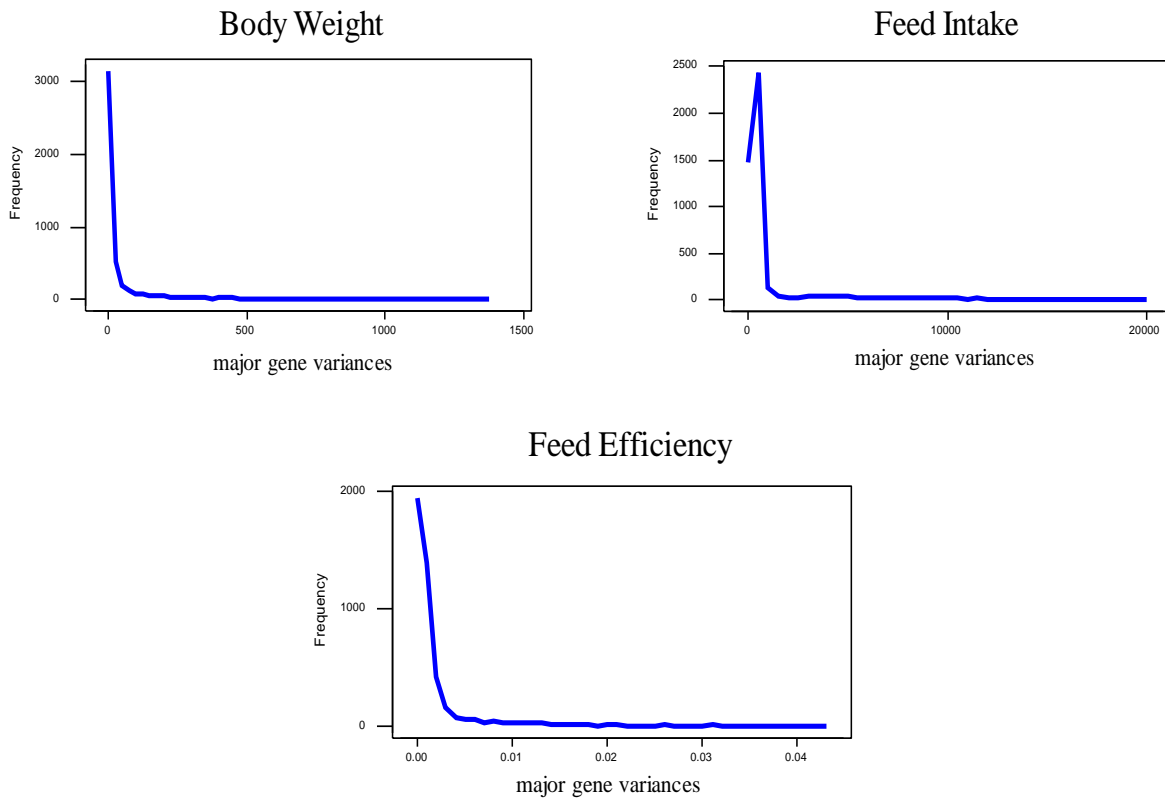**Figure 1.** Marginal posterior distributions of major gene variances for body weight, feed intake and feed efficiency.

Body Weight

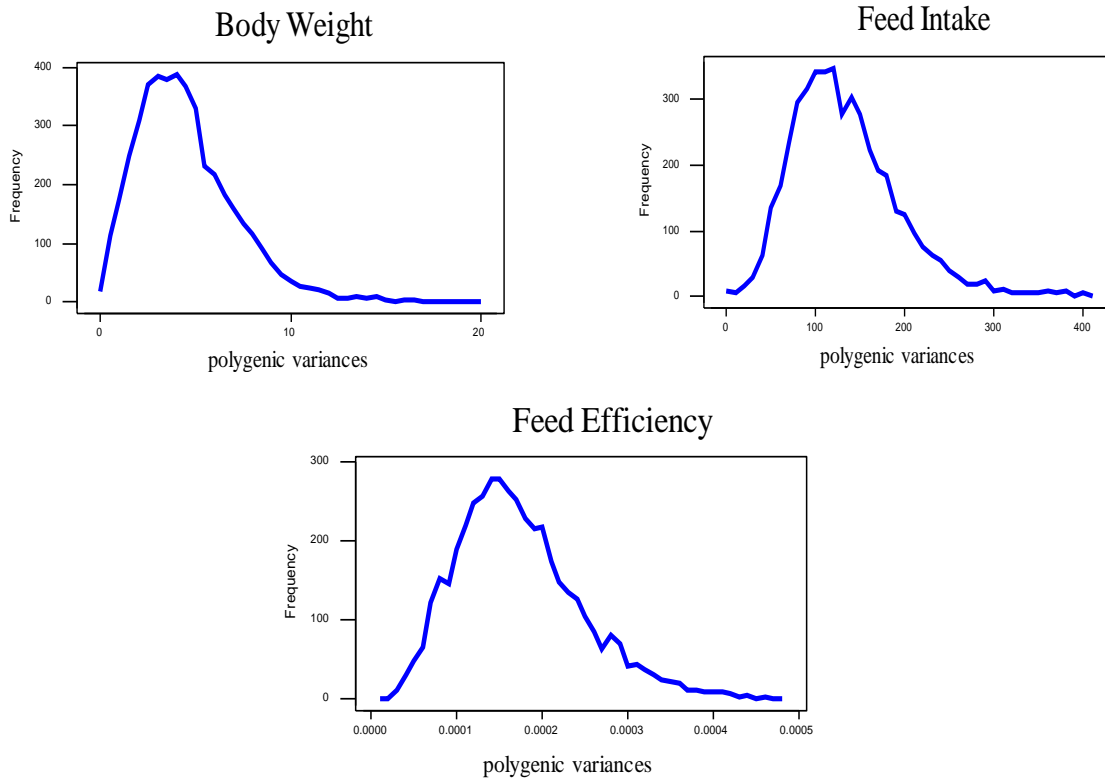## Feed Intake

Feed Efficiency

Figure 2. Marginal posterior distributions of polygenic variances for body weight, feed intake and feed efficiency.

Table 2. Posterior mean and standard deviation and left and right bounds of the 95 % highest posterior density for feed efficiency.

	Feed efficiency			
Parameter	**Posterior Mean**	**Posterior Standard Deviation**	**HPDR 95 % left**	**HPDR 95 % right**
Error variance	0.00017	0.00004	0.00000	0.00031
Polygenic variance	0.00017	0.00007	0.00001	0.00050
Major gene variance	0.00221	0.00502	0.00000	0.04408
Additive effect	0.05073	0.05327	0.00000	0.29994
Dominant effect	- 0.00347	0.05787	- 0.17884	0.31596

The results of segregation analyses of feed intake are given in Table 1. Polygenic variance (132.29) was found to be smaller than the major gene variance (919.99) (Figure 1 and Figure 2). Existence of the major gene for feed intake is confirmed by the 95 % HPDR, not including zero (Table 1). Dominance effect (24.88) was found to be nearly identical with the additive effect (24.80). Results of 95% HPDR for Mendelian transmission probabilities are given in Table 2. Since 95% HPDR is not included Mendelian transmission probabilities of 1, 0.5 and 0 it was concluded that mode of inheritance of feed intake is not Mendelian. Estimates of major gene heritability for feed intake were found to be around 0.96 while polygenic heritability found to be 0.35. Convergence analyses of the Gibbs sampler based on $\hat{R}$ given at Table 4. Convergence was concluded since estimates of $\hat{R}$ found to be around 1 for all parameters.

The results of segregation analyses of feed efficiency are given in Table 2. Polygenic variance (0.00017) was found to be smaller than the major gene variance (0.00221) (Figure 1 and Figure 2). Since 95 % HPDR included zero (Table 2) null hypothesis cannot be rejected, hence concluded that there is no segregation of major gene for feed efficiency. Dominance effect (- 0.00347) was found to be smaller than the additive effect (0.05073). Results of 95 % HPDR for Mendelian transmission probabilities are given in Table 2. Since 95 % HPDR is not included Mendelian transmission probabilities of 1, 0.5 and 0 it was concluded that mode of inheritance of feed efficiency is not Mendelian. Estimates of major gene heritability for feed efficiency were found to be around 0.93 while polygenic heritability found to be 0.52. Convergence analyses of the Gibbs sampler based on $\hat{R}$ given at Table 4. Convergence was concluded since estimates of $\hat{R}$ found to be around 1 for all parameters.

Table 3. Left and right bounds of 95 % highest posterior density regions for transmission probabilities, presented as the probabilities to inherit an A allele from AA, AB, and BB genotypes.

Body weight	Transmission probability	HPD 95 % left	HPD 95 % right
	Pr(A|AA)	0.00000	1.00323
	Pr(A|AB)	0.00000	1.00381
	Pr(A|BB)	0.00000	1.00335
Feed intake	Transmission probability	HPD 95 % left	HPD 95 % right
	Pr(A|AA)	0.00000	1.00271
	Pr(A|AB)	0.00000	0.96765
	Pr(A|BB)	0.00000	0.31410
Feed efficiency	Transmission probability	HPD 95 % left	HPD 95 % right
	Pr(A|AA)	0.00197	1.00200
	Pr(A|AB)	0.00200	1.00201
	Pr(A|BB)	0.00199	1.00383

Table 4. Investigation on the existence of convergence using $\hat{R}$ for body weight, feed intake and feed efficiency.

Parameter	Body weight	Feed intake	Feed efficiency
Error variance	1.000033	1.000108	**
Polygenic variance	1.000098	1.000084	**
Major gene variance	1.000923	0.99999	1.001863
Additive effect	1.002559	1.008708	1.008368
Dominant effect	1.000464	1.00132	1.010733

** There is convergence but the program could not calculate the number because it is too small.

Discussion and Conclusion

Animal breeding employs deep and complex pedigrees in breeding programs. Estimating the breeding value with high precision allows higher genetic improvement (11). Similarly, the use of deep and complex pedigrees in the analysis of major genes allows more accurate predictions and higher genetic improvements (7, 16). Gibbs sampling allows accurate statistical interpretation from deep pedigrees for segregation analyses in animal science (9).

In this study, segregation analyses were performed using an F2 mouse dataset for body weight, feed intake and feed efficiency. Ehsani et al. (3) and Karacaören (12) reported a number of genes in association with the body weight, feed intake and feed efficiency by genome wide association analyses using the same mouse dataset. Ehsani et al. (3) have shown that three QTLs are effecting body weight. In this study, we also detected a strong major gene variance (Table 1 and Table 2) for body weight but using only pedigrees and phenotype information. However, the mode of inheritance of this major gene is not confirmed to be Mendelian (Table 3). Linkage disequilibrium (LD) may cause deviation from Mendelian inheritance. Since the genome wide association studies (GWAS) exploits the LD over the genome such a result (Table 3) also expectable. Deviation from Mendelian segregation ratios could be explained by physical closeness with genes under selection pressure, mating among relatives and founder effects (13, 14,). In addition to LD, results of Ehsani et al. (3) and Karacaören (12) showed that the assumption of just one major gene for body weight is not a correct assumption.

Sanchez et al. (15) also compared the results of QTL (using interval mapping) and segregation analyses in pigs and reported that the at least presence of two QTL's could be responsible for the partial disagreement between the two approaches.

Jarvik (10) reported the relationship between normality and interpretation of the results of segregation analyzes. We applied various logarithmic and geometric transformations to obtain normality for body weight (results are not shown) but the major gene component was found to be non-significant after the transformations. Sanchez et al. (15) also reported decreased power due to

the Box-Cox transformations for the segregation analyses using a pig population.

Wolc et al. (16) found the heritability for body weight at the range of 0.25 to 0.47 using number of generations in mice with a polygenic model, whereas we found 0.42. Estimated polygenic heritabilites for body weight (h^2_{BW}=0.42), feed intake (h^2_{FI}=0.53) and feed efficiency (h^2_{FE}=0.58) was proportionally higher those obtained by Ehsani et al. (3): as h^2_{BW} =0.29, h^2_{FI} =0.35 and h^2_{FE} =0.52 respectively. Similar to our results Sanchez et al. (15) and Wolc et al. (16) reported that the results of segregation analysis overlap with the results of DNA analysis.

The present study was designed to determine if the segregation analyses and the genomic analyses (3) gave similar or contrasting results. The most striking result to emerge from the data (Table 1) is that existence of a major gene for body weight and feed intake was confirmed by both the segregation analyses and GWAS (12). Contrary to expectations this study did not detect any evidence for the Mendelian inheritance of the body weight or feed intake (Table 3). These results reflect those of (3) who also found that QTLs with major (large) and small effects at various chromosomes. In accordance with the present results, previous studies have demonstrated (2, 5) the importance of the genetic architecture of the phenotypes.

The present study provides the first comprehensive empirical assessment of genomic and segregation analyses under laboratory (controlled environment and homogenized genetic material) conditions for body weight, feed intake and feed efficiency in mice. The scope of this study was limited in terms of genetic and environmental conditions. Further studies need to be carried out in order to validate the findings of the current study at the farm level. These findings suggest (confirm) a role for segregation analyses for detection of a major gene(s) when the DNA information is not available due to economic reasons.

Acknowledgement

Conflict of Interest

The authors are declared that there is no conflict of interest.

References

1. **Allan MF, Pomp D, Eisen EJ** (2005): *Genomic mapping of direct and correlated responses to long-term selection for rapid growth rate in mice.* Genetics, **170**, 1863-1877.
2. **Barton NH, Etheridge AM, Véber A** (2017): *The infinitesimal model: definition, derivation, and implications.* Theor Popul Biol, **118**, 50-73.
3. **Ehsani A, Sørensen P, Pomp D, et al** (2012): *Inferring genetic architecture of complex traits using Bayesian integrative analysis of genome and transcriptome data.* BMC Genomics, **13**, 456.
4. **Falconer DS, Mackay TFC** (1996): *Introduction to Quantitative Genetics.* Longman Group Ltd, UK.
5. **Flint J, Mackay TF** (2009): *Genetic architecture of quantitative traits in mice, flies, and humans.* Genome Res, **19**, 723-733.
6. **Gelman A, Carlin JB, Stern HS, et al** (2004): *Bayesian Data Analysis.* Chapman and Hall, USA.
7. **Janss LLG, van Arendonk JV, Brascamp EW** (1997): *Bayesian statistical analyses for presence of single genes affecting meat quality traits in a crossed pig population.* Genetics, **145**, 395-408.
8. **Janss LLG** (2008): *iBay manual version 1.46.* Janss Bioinformatics, Lieden, Netherlands.
9. **Janss LLG** (1996): *Statistical identification of major genes in pigs. PhD Thesis.* Wageningen Agricultural University, The Netherlands.
10. **Jarvik GP** (1998): *Complex segregation analyses: uses and limitations.* Am J Hum Genet, **63**, 942–946.
11. **Karacaören B, Fırat MZ** (2012): *Genetik ilerlemelerin hesaplanmasında kullanılan istatistiksel yöntemlerin karşılaştırılması.* Ankara Univ Vet Fak Derg, **59**, 115-120.
12. **Karacaören B** (2014): *Empirical comparison of association and admixture mapping for body weight using F2 mice data set.* Archiv Tierzucht, **57**, 1-6.
13. **Medina-Macedo L, Lacerda AEB, Ribeiro JZ, et al.** (2014): *Investigating the Mendelian inheritance, genetic linkage, and genotypic disequilibrium for ten microsatellite loci of Araucaria angustifolia.* Silvae Genet, **63,** 234-239.
14. **Nielsen R, Slatkin M** (2013): *An Introduction to Population Genetics: Theory and Applications.* Sinauer Associates is an imprint of Oxford University Press, UK.
15. **Sanchez MP, Iannuccelli N, Basso B, et al** (2007): *Identification of QTL with effects on intramuscular fat content and fatty acid composition in a Duroc × Large White cross.* BMC Genet, **8**, 55.
16. **Wolc A, Skotarczak E, Schlote W, et al** (2009): *Single-gene Effects on Body Weight in Selected and Unselected Mouse Lines Detected by Bayesian Marker-free Segregation Analysis.* Scand J Lab Anim Sci, **36,** 185-191.