APPLICATION OF RANDOM REGRESSION TECHNIQUES TO DISSECT AGE-DEPENDENT QUANTITATIVE TRAIT LOCI FOR GROWTH IN LAMBS

G. Hadjipavlou[1,2] and S.C. Bishop[1]

[1]Roslin Institute and R(D)SVS, University of Edinburgh, Roslin, Midlothian EH25 9PS, UK
[2]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

SUMMARY

A systematic procedure for selecting random regression (RR) models that include quantitative trait loci (QTL) as time-dependent random effects, in addition to polygenic and environmental effects, was investigated using live weight data measured at 4-week intervals on lambs from 9 half-sib families of Scottish Blackface sheep. Microsatellite marker information was used to calculate gametic relationship matrices. The RR model allowed the detection of apparently significant QTL on chromosomes 14 and 20 when a RR polynomial of 2nd order or higher was fitted for the QTL effect. However, generally the choice of RR models with relevant variance components and the appropriate order for the random coefficient parametric curves proved complex for models both with and without QTL. With the given data structure, the RR approach had poorer ability to describe growth QTL over time compared to a growth model approach previously used to dissect the expression of age-dependent QTL for growth rates and live weights.

INTRODUCTION

Growth is a longitudinal trait that is a composite of growth rate phenotypes over time. Patterns of genetic correlations among live weights at different ages often demonstrate the trait complexity. For example, using sheep data Riggio *et al.* (2008) showed that inter-age genetic correlations for live weight, whilst strongly positive, are often less than unity, with the correlation decreasing as the time between weight measurements increases. Thus, it is likely that distinct loci act on live weights at different growth stages. For the detection of quantitative trait loci (QTL) that are associated with growth or live weight, it would be beneficial to simultaneously analyze multiple measurements and take account of the correlation structure among measurements ordered across time. Random regression (RR) methodology, in principle, provides a means of longitudinal trait analysis, while accounting for the covariance structure among measurements and allowing flexible model fitting. The main objective of this study was to examine the steps and issues relating to the use of RR to dissect QTL and to utilise RR models for chromosome-wide detection and quantification of QTL that influence longitudinal live weights, using data from Scottish Blackface sheep. We also sought to compare the results from applying RR for detecting age-dependent QTL with those of an alternative technique used previously for growth QTL mapping by analysing descriptors of growth derived from fitting a growth curve (Hadjipavlou and Bishop 2009).

MATERIALS AND METHODS

Data. The data comprised live weights at birth and at 4-week intervals up to 24 weeks, and genotypes for microsatellite markers on OAR 1-3, 5, 14, 18, 20, and 21, for 788 lambs from 9 half-sib families of Scottish Blackface sheep. A total of 135 markers were used (range=8-34, median= 15). Number of progeny per family ranged from 34 to 154 individuals, bred over 3 years (2001–2003). Standard records (parentage, day of birth, sex, rearing type) were collected. The pedigree comprised 1119 animals, including two generations of sire and dam ancestors.

Random Regression (RR) model. A RR model was used to fit the longitudinal live weight data first without accounting for a QTL and subsequently with the inclusion of a QTL effect. The generalized RR equation used for the full model is: $\mathbf{y= X\beta + Wq + Z_1u + Z_2p + Z_3l + Z_4e}$
where **y** is a vector of observations taken at several time points for each lamb; **β** is a vector of age-dependent fixed effects; **X** is the design matrix connecting fixed effects with records; $\mathbf{Wq}$, $\mathbf{Z_1u}$, $\mathbf{Z_2p}$, $\mathbf{Z_3l}$, $\mathbf{Z_4e}$ are the systematic time-dependent deviations from the fixed curve, modelled as random effects, due to allelic effects of the QTL, polygenic, permanent animal, litter (common environment) and residual effects, respectively. Vector **q** is of dimensions $2N_gp_1$, where N_g is the number of animals included in the gametic matrix. Vector **u** is of dimension N_ap_2, N_a being the number of animals in the relationship matrix (i.e. pedigree). The permanent animal vector **p**, the common environment vector **l** and the residual vector **e** are of dimensions N_pp_3, N_pp_4 and N_pp_5, respectively, where N_p is the number of animals with records. Parameters p_1, p_2, p_3, p_4, p_5 correspond to the number of random regression coefficients used to model the associated random effect. Matrices **W**, $\mathbf{Z_1}$ and $\mathbf{Z_2}$, $\mathbf{Z_3}$ are design matrices of the RR curve, including covariates. The elements of these matrices are $\Phi_i=\forall_i(t^*_{ij})$, where $\forall_i$ are coefficients of the chosen Legendre polynomial for lamb i at age j (t^*_{ij}). The age values are standardized between -1 and +1. Because heterogeneous residual variance was assumed, $\mathbf{Z_4}$ was fitted as a diagonal matrix of distinct variance for each age class. The random vectors **q**, **u**, **p**, **l**, **e**, are assumed to be mutually independent and to follow multivariate normal distributions; $\mathbf{q_i}|M, c_i \sim MVN(0,\mathbf{K_{0i}}\otimes\mathbf{Q_i}|M,c_i)$, $\mathbf{u} \sim MVN(0,\mathbf{G_0}\otimes\mathbf{A})$, $\mathbf{p} \sim MVN(0, \mathbf{P_0}\otimes\mathbf{I})$, $\mathbf{l} \sim MVN(0, \mathbf{L_0}\otimes\mathbf{I})$, $\mathbf{e} \sim MVN(0, \mathbf{I\sigma^2_{ek}})$ where $\mathbf{K_{0i}}$, $\mathbf{G_0}$, $\mathbf{P_0}$ and $\mathbf{L_0}$ are covariance matrices among random regression coefficients. Matrix **A** is the additive genetic relationship matrix and $\mathbf{Q_i}|M,c_i$ is the gametic relationship matrix of the allelic effects at the QTL, conditional on marker data (M) and the position (c_i) on the chromosome.

Model Choice and Statistical Testing. Prior to fitting a QTL, the Likelihood Ratio (LR) test was used to assess the significance of polygenic, permanent animal and litter effects across nested mixed RR models with varying order of the random regression polynomials for each effect. In all models, relevant identifiable fixed effects and all two-way interactions of month with these fixed effects were fitted. A fixed regression of 5th order, analogous to the population mean in single time-point analyses, and heterogeneous residual variances for the 7 monthly intervals were also fitted. The optimal RR model was then used to fit a QTL effect for each chromosome. The QTL effect was also modelled with random polynomial curves to allow for systematic effects of the QTL on the deviation of the animal phenotype from the expected value over time, and hence allowed for changes in the QTL variance with age. The gametic relationship matrix was calculated across all animals using Loki (Heath 1997), and it was utilised in the model to include the allelic effects of the QTL conditional on marker data and the position on the chromosome. LR tests were used to assess the significance of QTL with different order for the random regression polynomial on chromosomes 14 and 20. All RR model analyses were performed using ASReml.

RESULTS AND DISCUSSION

No-QTL RR model choice. Table 1 shows the log likelihood (LogL) estimates from a group of models in which a polygenic effect was fitted with a Legendre polynomial of order 1. Nested LR testing can only be performed across rows or down columns, and the df associated with each order of fit are 1, 3, 6, 10, 15 and 21, respectively. Because RR describes individual deviations from the average curve of the population, a 2nd order RR polynomial can be significant when orders 0 or 1 are not. Some models did not converge. In other cases in which the LogL did not improve when the order of either litter or permanent animal effect was increased, the higher order model was rejected as the likelihood maximisation process had probably failed to reach a global maximum.

Based on the above, all models with either of the two environmental effects fitted to order higher than 3 were rejected. From the remaining models, a model with a 2nd order RR polynomial fitted for both environmental effects was chosen to reduce the possibility of model overfitting. Inter-age genetic correlations estimated from this model were positive yet decreased as the interval between weight measurements increased, and were in agreement with those reported by Riggio *et al.* (2008).

Table 1 Log-likelihood estimates for no QTL RR model fitting live weights over time

Random regression polynomial order[1,2]		Permanent animal effect					
		0	1	2	3	4	5
Litter effect	0	-6299.87	-6335.83	-6131.82	-6130.64	NC[3]	NC
	1	-6302.58	-6393.58	-6192.8	-6194.48	NC	NC
	2	-6118.01	-6224.90	-6189.98	-6163.19	-6159.74	-6180.33
	3	-6130.09	-6243.11	-6178.88	-6194.03	-6201.74	NC
	4	-6172.57	-6286.41	-6177.15	-6199.01	-6231.62	NC
	5	-6219.32	-6332.14	-6195.83	-6211.04	-6249.87	NC

[1]A polygenic effect with a random regression polynomial of order 1 was fitted in each model.
[2]RR polynomial order is equivalent to order+1 random coefficients estimated.
[3]NC=Model fit did not converge.

QTL RR model choice and estimates. Chromosomes (OAR) 14 and 20 were explored as age-dependent QTL were previously found on these chromosomes, using live weights predicted from growth curves (Hadjipavlou & Bishop 2009). For each chromosome and each order of RR for the QTL effect, the likelihood of the full model was maximised every 5cM. Formal model testing included LR test and a conservative chi-squared test statistic were employed to assess the overall QTL significance, as well as the significance of increasing the order of the RR polynomial fitted for the QTL, at the chromosome position with the highest likelihood. The LogL and test statistics for an OAR 14 QTL are shown in Table 2. A first order Legendre polynomial for the QTL effect resulted in a model for which the QTL was not significant. Additionally, the QTL variance was not estimable (or estimated to be zero) at certain chromosomal positions. A significant QTL was supported by models with polynomial order of 2 or higher. Further, the statistic for increasing the order of the RR polynomial provided justification for fitting a QTL with a 3rd order RR.

Table 2 Log-likelihood estimates and test statistics for RR model with chromosome 14 QTL effect fitted using random coefficient polynomials

QTL RR order in model	LogL	Test statistic when compared to no QTL model[1]	Test statistic of increasing QTL order[1]
1	-6188.91	2.14 (3 d. f.)	NA
2	-6183.47	13.02 (6 d. f.)*	10.88 (3 d. f.)**
3	-6158.9	62.16 (10 d. f.)**	49.14 (4 d. f.)**
4	-6156.43	67.10 (15 d. f.)**	4.94 (5 d. f.)

[1]QTL effect: *significant (P<0.05); ** highly significant (P<0.01)

Figure 1 shows the proportion of phenotypic variance partitioned to the QTL effect across time in a model with a QTL RR of order 3. Since the estimated variance ratios across time differ only marginally, it is difficult to come to a conclusion regarding the trend in variance change across time. A time-dependence trend may be speculated, with a maximum effect around 60 days of age, but it is masked further by inflated variance estimates at the end age points, which arise as a

consequence of prediction using polynomial regression. A similar pattern of results was seen for OAR 20. Our results also showed confounding of the permanent animal and QTL effect, presumably because both terms are correlated with an animal's Mendelian sampling term.

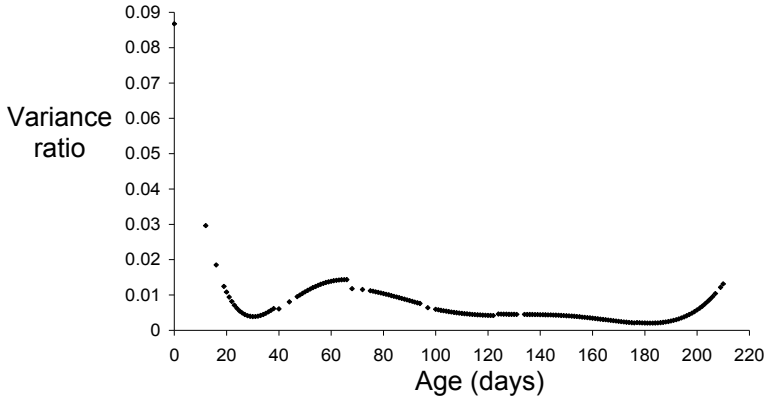

Figure 1. Proportion of phenotypic variance explained by the QTL, across age, on OAR 14.

Comparison between growth model and RR approaches and results. We have previously demonstrated the use of growth models to dissect the expression of age-dependent QTL. An OAR 14 QTL, which was only significant around 8 weeks of age, was detected and an OAR 20 QTL displayed a significance trajectory across age, and was significant from 8-16 weeks, with maximum significance at 12 weeks. The RR results support the above findings since on both chromosomes a QTL effect of order 2 or higher was deemed significant in a model in which polygenic and environmental effects were also modelled using RR. The fact that higher polynomial orders proved necessary to identify the QTL suggests a complex time dependence of QTL effects, as assessed by deviations from the expected age-dependent live weight. The pattern of QTL variance over time remains to be resolved, as results from our analysis are ambiguous.

CONCLUSIONS

Random regression techniques have been previously proposed for dissecting time-dependent polygenic and environmental effects, and could potentially be used to identify and describe the expression of age- or time-dependent QTL. Our study explored a systematic procedure for selecting random regression models that included a QTL as a time-dependent random effect, in addition to polygenic and environmental effects. High order RR models apparently detect time-dependent QTL, even though the pattern and magnitude of the change in QTL variance across time remains unclear. Using simulations, future work will assess the power of RR models for detection of age-dependent QTL, and explore the data requirements for such analyses.

ACKNOWLEDGEMENTS

This work was conducted as part of the GENACT Project, funded by the Marie Curie Host Fellowships for Early Stage Research Training, as part of the 6th Framework Programme of the European Commission. Defra funded the resource flock and genotyping.

REFERENCES

Hadjipavlou, G. and Bishop, S.C. (2009) *Anim. Genet.* **40**:165.
Heath S. (1997) *Am. J. Hum. Genet*. **61**:748.
Riggio, V., Finocchiaro, R. and Bishop S.C. (2008) *J. Anim. Sci*. **86**:1758.