Bodyweight QTL on mouse chromosomes 4 and 11 by selective genotyping: regression v. maximum likelihood

Beben Benyamin A, B, D, E, Ian C. A. Martin A, Carol C. Cheung C, Michael F. Buckley C, Peter C. Thomson A, Peter M. Visscher D and Chris Moran A

A Centre for Advanced Technologies in Animal Genetics and Reproduction (Reprogen), University of Sydney, NSW 2006, Australia.
B Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JT, UK.
C Molecular and Cytogenetics Unit, South Eastern Area Laboratory Services, Department of Haematology, Prince of Wales Hospital, Sydney, NSW 2031, Australia.
D Genetic Epidemiology, Queensland Institute of Medical Research, 300 Herston Road, Brisbane, Qld 4029, Australia.
E Corresponding author. Email: bebenB@qimr.edu.au

Abstract. Two quantitative trait loci (QTL) for mouse bodyweight have previously been reported on mouse chromosomes (MMU) 4 and 11 from crosses of a highly fecund and large mouse strain, Inbred Quackenbush-Swiss 5 (QSi5) and C57Bl/6J. QSi5 has now been crossed with CBA/CaH to produce 1128 F2 mice to confirm the existence and effect of these QTL. In total, 226 mice from the upper and lower deciles for bodyweight were genotyped using 12 microsatellite markers covering MMU4 and MMU11. Regression and maximum likelihood based interval mapping by either including all mice (un-genotyped mice were treated as having missing genotypes) or including only selectively genotyped mice in the analyses were used to estimate the positions and effects of the QTL. The results confirmed the existence and effects of both QTL. Although all methods estimated the same QTL positions, the QTL effects were overestimated compared with the estimates using a suggested method (maximum likelihood by including all mice in the analysis). However, the overestimated QTL effects could be mathematically corrected. Since the confidence intervals of both QTL are still too large for positional cloning, an advanced intercross line is being bred for finely mapping these bodyweight QTL.

Introduction

Mouse bodyweight has been recognised as an excellent model for studying the genetic architecture of quantitative traits, important to both human health, mainly obesity, and growth rate in domestic animals. Genetic variation in bodyweight, as for any other quantitative trait, is controlled by many genes, of which the effects of some are sufficiently large that the quantitative trait loci (QTL) responsible can be linkage-mapped with the aid of genetic markers, such as microsatellites and single nucleotide polymorphisms.

A major objective of mapping QTL for bodyweight in mice is to estimate the number and position of loci affecting the trait, the magnitude of their effects, and type of gene action (Corva and Medrano 2001). Another objective is to identify the genes underlying the QTL, so that information can be transferred from the model to human health and livestock productivity studies.

Numerous QTL for bodyweight or growth-related traits and obesity in mice have been reported in recent years. Different strains of mice and resource pedigrees (F2 or backcross) have been used to localise the QTL, with Brockmann and Bevova (2002) reviewing 85 QTL for bodyweight and 75 QTL for obesity. With the exception of the Y chromosome, at least one such QTL has been mapped on all of the mouse chromosomes with a concentration on MMU1, MMU7, and MMU11.

Using an F2 and backcross population derived from a large and highly fecund inbred mouse, Inbred Quackenbush-Swiss line 5 (QSi5) developed at the University of Sydney (Holt et al. 2004) and a standard inbred mouse, C57Bl/6J, Kirkpatrick et al. (1998) reported two QTL for 6-week bodyweight on MMU4 and MMU11 at 62.5 and 23 cM from the centromere, respectively. The QTL alleles that originated from QSi5 were found to increase the bodyweight. In the present study, an F2 intercross between QSi5 and another standard inbred mouse, CBA/CaH, has been utilised to confirm both the existence and effects of these QTL before attempting fine mapping. This confirmation is very important since false positives could occur (Falconer and Mackay 1996).

Selective genotyping is an efficient method of QTL mapping, where only individuals that show an extreme phenotype of interest are genotyped (e.g. Lander and Botstein 1989; Darvasi and Soller 1992; Darvasi 1997). It substantially reduces the number of individuals genotyped with little loss in statistical power (Lander and Botstein 1989; Xu and Vogl 2000).
Although regression interval mapping is computationally fast, Lander and Botstein (1989) suggested that it cannot be used to analyse selectively genotyped data. Instead, maximum likelihood interval mapping with inclusion of phenotypic data of the ungenotyped animals (treating unscored genotypes as missing values) should be used to map QTL from such data (Lander and Botstein 1989). Henshall and Goddard (1999) pointed out that regression interval mapping from selectively genotyped data will produce a bias in parameter estimation. In particular, the phenotypic effect will be overestimated due to selection bias (Lander and Botstein 1989). However, Knott (2005) argued that although QTL effects will be overestimated, regression interval mapping by analysing only selectively genotyped individuals will not bias the QTL location or cause spurious QTL to be detected. If the ungenotyped individuals were included in the regression analysis and treated as individuals with missing genotypes, then test statistics will be inflated as well. Therefore, besides the confirmation of bodyweight QTL in a different cross, another aim of the present study is the comparison of regression vs. maximum likelihood interval mapping and estimation of the magnitude of bias of regression methods applied to a selectively genotyped experimental dataset.

Materials and methods

This study was performed under Animal Care and Research Ethics approval N02/2–2001/2/3337 from the University of Sydney.

Mice breeding and husbandry

F₂ mice (1128 total) were generated from intermating between F₁ of a cross between a QSi5 female and a CBA/CaH male. QSi5 is an inbred line with a very large mature bodyweight and very high fecundity (Holt et al. 2004). CBA/CaH, a standard inbred line with more typical bodyweight was obtained from the Animal Resource Centre, Perth, Western Australia. QSi5 and CBA/CaH bodyweights differ by at least six phenotypic standard deviations. A comparison of 8-week bodyweight between parental lines (QSi5 and CBA) and F₂ mice is presented in Table 1. As expected, if there were genes affecting bodyweight segregating in the F₂ mice, the standard deviation of bodyweight for F₂ mice is higher than that of the two parental lines. It is also noted that the bodyweight of QSi5 females was more variable than that of males. However, since the sample sizes of the two parental lines were small compared with that of F₂ mice, this difference could just be a fluctuation due to sampling.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Sex</th>
<th>n</th>
<th>Bodyweight (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QSi5</td>
<td>Male</td>
<td>6</td>
<td>31.6 ± 1.4</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>21</td>
<td>26.7 ± 2.4</td>
</tr>
<tr>
<td>CBA/CaH</td>
<td>Male</td>
<td>20</td>
<td>18.7 ± 1.3</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>17</td>
<td>15.5 ± 1.0</td>
</tr>
<tr>
<td>F₂</td>
<td>Male</td>
<td>561</td>
<td>30.6 ± 2.7</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>567</td>
<td>24.7 ± 2.6</td>
</tr>
</tbody>
</table>

Bodyweight of F₂ mice was measured at 8 weeks of age using a Mettler PM6000 balance to 0.1 g accuracy. The spleen from each mouse was collected and stored in liquid nitrogen for DNA extraction. Thirty-five different collection days were required due to the limited number of mice that could be processed in 1 day for the platelet count project.

In order to identify extreme mice for selective genotyping, the F₂ bodyweights were adjusted for fixed effects, including sex and collection days using a general linear model analysis and the residuals obtained. Both sex and collection days have highly significant effects on bodyweight, and the residuals were found to be normally distributed (Anderson-Darling \( A^2 = 0.70; P\text{-value} = 0.07 \)). Twenty percent (226 mice) of F₂ mice (10% from each tail of the distribution) were selected for selective genotyping.

DNA extraction and genotyping

Genomic DNA was extracted from the spleen samples using a standard procedure (Sambrook et al. 1989) with some modifications, consisting essentially of Proteinase-K digestion to remove proteins and RNase treatment to remove RNA, followed by phenol–chloroform extraction to remove residual protein. Twelve microsatellite markers obtained from Research Genetics (Huntsville, AL) (six from each chromosome) were selected to provide an average spacing of 15 cM. These markers are D4Mit264, D4Mit345, D4Mit178, D4Mit219, D4Mit203 and D4Mit357 from MMU4 and D11Mit71, D11Mit231, D11inds9, D11Mit325, D11Mit198 and D11Mit167 from MMU11. The positions and orders of these markers were obtained from the Mouse Genome Databases (MGD) (MGD 2002).

Microsatellite markers were amplified by polymerase chain reaction (PCR) with fluorescently labelled primers. 100 ng genomic DNA was amplified with 0.1 mM dye labelled primers, 0.2 mM dNTPs, 2 mM MgCl₂, 1.5 µL 10x PCR buffer (100 mM Tris.Cl, 15 mM MgCl₂, 500 mM KCl, pH 8.3), 1 unit Taq polymerase and 9.1 µL H₂O using PTC-100 Programmable Thermal Controller. The denatured PCR products were then genotyped using an ABI-373 Sequencer, with genotypes called automatically with Genotyper 1.1 (Applied Biosystems, Foster City, CA) software and then checked manually using Genescan 2.1 (Applied Biosystems, Foster City, CA).
**Interval mapping**

Interval mapping tests for the presence of a QTL within a marker interval using an estimated genetic map as the framework. The likely position of a QTL is determined by comparing the likelihood of the dataset given the presence of QTL at the test position (alternate hypothesis) vs. the likelihood with no QTL (null hypothesis) (e.g. Doerge 2002). While maximum likelihood uses all available information from the marker-trait distribution (observed marker genotypes and the trait distribution) for estimating the QTL position and effects, regression methods use only marker genotype means. When there is no selection of the data, both methods yield similar answers (Haley and Knott 1992).

For selectively genotyped data, maximum likelihood has been suggested to map QTL by including all individuals with phenotypic data in the analysis, and treating the ungenotyped individuals as having missing genotypes (Lander and Botstein 1989). The ungenotyped individuals are needed in order to correct the bias of parameter estimation (Xu and Vogl 2000). Although the regression method yields an upward bias of parameter estimation (Henshall and Goddard 1999), Johnson et al. (1999) showed that for equal proportions of individuals selectively genotyped in each tail, the regression method is still a useful and quick method to determine the location of a QTL and is comparable to the maximum likelihood method. Parameter estimates from regression can be corrected using a multiplicative factor based on a method suggested by Darvasi and Soller (1992), i.e. (1 + x × i), where x and i are the deviation of the truncation point from the mean and the mean of the selected group (in s.d. units) of the truncated normal distribution, respectively. For 10% selection of each tail, the correction factor for the regression results (the effects and test statistics) in the present study is (1 + 1.28 × 1.76) = 3.25.

The additive effect of a QTL was estimated as half of the difference between QSi5 and CBA genotypes as follows,

\[ a = \frac{\bar{y}_{QQ} - \bar{y}_{qq}}{2} \]

and the dominance effect was defined as

\[ d = \bar{y}_{Qq} - \frac{\bar{y}_{QQ} + \bar{y}_{qq}}{2} \]

where Q refers to the QSi5 allele, q refers to the CBA allele, and \( \bar{y}_{QQ}, \bar{y}_{qq}, \bar{y}_{Qq} \) are the (residual) bodyweight means for each QTL genotype.

In the present study, both maximum likelihood and regression interval mapping methods were assessed on selectively genotyped experimental data. In each case, the analysis included either all mice with phenotypic data in the analyses (treating unscored genotypes as missing values) or only selectively genotyped mice. Thus, in total four different analyses were performed, namely: (1) *ml-all*: maximum likelihood by including all mice in the analysis (the recommended method [Lander and Botstein, 1989]), (2) *reg-all*: regression by including all mice in the analysis, (3) *ml-select*: maximum likelihood by including only selectively genotyped mice and (4) *reg-select*: regression by including only selectively genotyped mice. Maximum likelihood interval mapping was carried out using Windows QTL Cartographer Version 2.5 (Wang et al. 2006), whereas, regression interval mapping was performed using QTL Express (Seaton et al. 2002).

The residuals of bodyweight after adjusting for the effects of sex and collection days were used in all analyses. The 95% confidence interval (CI) of QTL position was determined using one LOD drop-off method, where the CI was calculated by finding the location on either side of QTL position that correspond to a decrease in the LOD score of 1 unit (Lander and Botstein 1989). However, it should be noted that the one LOD drop-off confidence interval is too anti-conservative, i.e. that fewer than 95% of one LOD intervals contain the QTL (Visscher et al. 1996).

In order to provide a comparable result between regression and maximum likelihood methods, F-statistics obtained from the regression method were converted to a likelihood ratio test (LRT) statistic using the following approximation:

\[ LRT = n \ln \left( 1 + \frac{2}{n-1} F \right) \]

with n the sample size (following Haley and Knott 1992).

**Meta-analysis**

A meta-analysis (Khatkar et al. 2004) and a test of the heterogeneity of QTL position in MMU4 and 11 were performed in order to evaluate whether the reported QTL from multiple studies (including the present study) are consistent with a single QTL in the region. The heterogeneity of QTL position test was performed by comparing the sum of the maximum LRT scores from the individual studies with the maximum LRT score obtained from adding up the profile LRT curves over the length of the chromosome. Under the null hypothesis that there is single QTL causing the reported effects in all studies, the difference between these LRT scores is distributed as a chi-square with n-1 degrees of freedom, where n is the number of studies.

**Results**

Since three microsatellite markers were either monomorphic or could not be scored, genotype data from nine markers only were included in the analyses (five and four markers on MMU4 and MMU11, respectively). The plots of the most likely QTL position obtained from both maximum likelihood and regression interval mappings are shown in Fig. 1. It can be seen from the plots that all methods estimated the same positions of the QTL. On MMU4, a bodyweight QTL was estimated to be at 48 cm from the centromere. The second peak at 68 cm did not reach chromosome-wide significance. On MMU11, a bodyweight QTL was also observed at 10 cm from the centromere. LRT statistic for both QTL and using all microsatellite markers was obtained by 1000 permutation tests. Except for *reg-all*, the LRT statistic profiles from the other methods were very close to the expected correction value, i.e. 3.25. As
estimated using maximum likelihood method, these QTL explain only a small proportion of the phenotypic variance (2.4% and 2.7% of the phenotypic variance were explained by QTL on MMU4 and MMU11, respectively).

The likelihood ratio test statistics obtained from all methods were all very similar, except for reg-all, which is around three times bigger. A complete comparison of parameter estimations between maximum likelihood and regression methods is presented in Table 2.

Fifteen studies (including Kirkpatrick et al. 1998) reporting QTL for bodyweight on either MMU4 or MMU11 were reviewed in the present study. Since only five studies contained sufficient information for the tests (test statistic profiles or standard error of QTL positions), i.e. three studies on MMU4 (i.e. Suto et al. 1998; Morris et al. 1999; Plum et al. 2000) and two studies on MMU11 (i.e. Hirayama et al. 1999; Moody et al. 1999), both meta-analysis and the heterogeneity of QTL position test were only applied to those studies and the present study. The results showed evidence for the existence of more than one QTL on both chromosomes, particularly for MMU4. However, in a closer examination of 13 studies reporting QTL for late growth bodyweight on MMU4, five studies (i.e. Brockmann et al. 1998, 2000; Kirkpatrick et al. 1998; Plum et al. 2000; Rocha et al. 2004) identified a similar QTL position, i.e. between 50 and 60 cM. In contrast, on MMU11, there was no obvious consensus position of any QTL.

**Discussion**

Two bodyweight QTL with alleles originating from QSi5 increasing bodyweight in mice have been reported by Kirkpatrick et al. (1998). Using a large number of F2 mice derived from a cross of CBA/CaH with QSi5, both the positions

---

![Graph showing likelihood ratio test statistics](image)

**Fig. 1.** The peak in likelihood ratio test (LRT) statistic at 48 and 10 cM from the centromere shows evidence of a QTL for bodyweight on (a) MMU4 and (b) MMU11, respectively. reg-all: regression by including all mice in the analysis (dashed line); ml-all: maximum likelihood by including all mice in the analysis (solid line); reg-select: regression by including only selectively genotyped mice (dotted line); ml-select: maximum likelihood by including only selectively genotyped mice (dash-dot line).

**Table 2.** The comparison of test statistics and parameter estimations obtained from regression and maximum likelihood methods using either all mice or only selectively genotyped mice in the analyses

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ml-all</th>
<th>reg-all</th>
<th>ml-select</th>
<th>reg-select</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MMU 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position (cM)</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>95% confidence interval (CI) (cM)</td>
<td>28–55</td>
<td>40–52</td>
<td>26–54</td>
<td>28–54</td>
</tr>
<tr>
<td>Additive effect (g)</td>
<td>0.56</td>
<td>1.84</td>
<td>1.90</td>
<td>1.86</td>
</tr>
<tr>
<td>Dominance effect (g)</td>
<td>-0.10</td>
<td>-0.30</td>
<td>-0.24</td>
<td>-0.31</td>
</tr>
<tr>
<td>Proportion of variance</td>
<td>2.4%</td>
<td>28%</td>
<td>8.3%</td>
<td>8.3%</td>
</tr>
<tr>
<td>Likelihood ratio test (LRT) statistic</td>
<td>13.24</td>
<td>47.25</td>
<td>14.29</td>
<td>14.37</td>
</tr>
<tr>
<td>LOD score</td>
<td>2.87</td>
<td>10.26</td>
<td>3.10</td>
<td>3.12</td>
</tr>
<tr>
<td>5% Empirical threshold (LOD)</td>
<td>2.23</td>
<td>2.07</td>
<td>2.71</td>
<td>2.16</td>
</tr>
<tr>
<td><strong>MMU 11</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position (cM)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>95% CI (cM)</td>
<td>0–25</td>
<td>0–17</td>
<td>0–24</td>
<td>0–24</td>
</tr>
<tr>
<td>Additive effect (g)</td>
<td>0.64</td>
<td>1.87</td>
<td>2.10</td>
<td>1.86</td>
</tr>
<tr>
<td>Dominance effect (g)</td>
<td>-0.27</td>
<td>-0.69</td>
<td>-0.62</td>
<td>-0.71</td>
</tr>
<tr>
<td>Proportion of variance</td>
<td>2.7%</td>
<td>30%</td>
<td>9.0%</td>
<td>9.0%</td>
</tr>
<tr>
<td>LRT statistic</td>
<td>12.60</td>
<td>42.97</td>
<td>12.75</td>
<td>12.69</td>
</tr>
<tr>
<td>LOD score</td>
<td>2.74</td>
<td>9.33</td>
<td>2.77</td>
<td>2.76</td>
</tr>
<tr>
<td>5% Empirical threshold (LOD)</td>
<td>2.06</td>
<td>1.94</td>
<td>2.04</td>
<td>1.96</td>
</tr>
</tbody>
</table>
and directions of effects of QTL alleles found in the present study are consistent with those reported by Kirkpatrick et al. (1998). Although bodyweight was measured in the present study at 8 weeks of age [2 weeks older than for Kirkpatrick et al. (1998)], Cheverud et al. (1996) suggested that growth after 6 weeks of age in the mouse is categorised as late growth. Thus, both Kirkpatrick et al. (1998) and the present study are examining the same late growth trait.

The QTL positions at both chromosom states in this study are around 10 cM closer to the centromere than those of Kirkpatrick et al. (1998), but their confidence intervals overlapped. The additive effects of both QTL are quite similar to those of Kirkpatrick et al. (1998), but their dominance effects are different in both sign and values. This difference could be due to the different strain used or less accurate estimation of dominance effects with selective genotyping.

The MGD (MGD 2002) lists five bodyweight/late growth QTL on MMU4, namely Bwg1 (6.3 cM); Wt41 (7 cM); Bw7 (59 cM); Pbwog2 (62 cM); Bw8q2 (66 cM) and seven on MMU11, namely Bglq8 (1.5 cM); Bw16 (14 cM); Wt10q3 (32 cM); Wt6q3 (36 cM); Wg4 (46 cM); Bw4 (55 cM); C10Bw3 (75 cM). The bodyweight QTL identified by Kirkpatrick et al. (1998) were not included in MGD, probably since the LOD scores (2.09 and 2.96 for QTL on MMU4 and MMU11, respectively) are only suggestive of linkage (Landers and Kruglyak 1995). The present study, however, supports the existence of QTL in similar chromosomal regions with similar statistical support.

Since the confidence intervals for the QTL are still large, further genetic mapping is required to refine the location of these QTL. An advanced intercross line (Darvasi and Soller 1992; Iraqi et al. 2000) derived from Q5S and CBA/CaH is being bred for fine mapping of platelet QTL (Cheung et al. 2004) and will also be used for fine mapping bodyweight QTL.

Selective genotyping employed in the present study substantially reduced the number of mice requiring genotyping. Both regression and maximum likelihood methods can be used to map the QTL from selectively genotyped datasets. All methods mapped the QTL into the same locations in both chromosomes. However, only maximum likelihood analysis that treated all individual (ml-all) provides an unbiased estimate of QTL effects [as suggested by Landers and Botstein (1989)]. The estimates of QTL effects from other methods were upwardly biased and should, therefore, be corrected using methods recommended by Darvasi and Soller (1992). The inflation of test statistic (LOD score) from regression method by including all mice (reg-all) compared with all methods has been noted previously by Knott (2005). While the maximum likelihood method properly used the information (i.e. used only phenotypic information) from ungenotyped individuals, the regression method calculated their expected genotypic values and treated these individuals as other individuals with genotypes. Thus, it increased the apparent sample size for calculating the test statistics, which in turn, inflated them.

Acknowledgements

We would like to thank Dr Kyall Zenger for the initial screening of the microsatellite markers. Beben Benyamin would like to thank AusAID for providing the Australian Development Scholarship for pursuing a Master of Agriculture degree at the University of Sydney, Australia.

References


Lander ES, Kruglyak L (1995) Genetic dissection of complex traits:
guidelines for interpreting and reporting linkage result. *Nature Genetics*
**11**, 241–247. doi:10.1038/ng1195-241
MGD (2002) Mouse Genome Database, Mouse Genome Informatics Web
Site, The Jackson Laboratory, Bar Harbor, Maine. Available at
http://www.informatics.jax.org [Verified 28 March 2007]
quantitative trait loci influencing traits related to energy balance in
growth traits in C57BL/6J × DBA/2J mice. *Mammalian Genome* **10**,
225–228. doi:10.1007/s003359900977
2 diabetes-like hyperglycemia in a backcross model of NZO and SJL
doi:10.1007/s00335-003-2312-x
manual.’ 2nd edn. (Cold Spring Harbour Laboratory Press: Cold Spring
Harbor, NY)
Express: mapping quantitative trait loci in simple and complex
bioinformatics/18.2.339
of obesity in KK mouse and effects of A¹ allele on quantitative
s003359900809
Department of Statistics, North Carolina State University, Raleigh, NC.
Available at http://statgen.ncsu.edu/qtlcart/WQTLCart.htm [Verified
4 April 2007]
under selective genotyping. *Heredity* **84**, 525–537. doi:10.1046/j.1365-
2540.2000.00653.x

Manuscript received 1 April 2006, accepted 18 October 2006