Genet. Res., Camb. (2005), **86**, *pp.* 209–221. *With* 4 *figures.* © 2005 Cambridge University Press
doi:10.1017/S0016672305007871 Printed in the United Kingdom

An application of Bayesian QTL mapping to early development in double haploid lines of rainbow trout including environmental effects

VICTOR MARTINEZ[1,2]*, GARY THORGAARD[3], BARRIE ROBISON[3]† AND MIKKO J. SILLANPÄÄ[4]

[1] *Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santa Rosa 11735, La Pintana, Santiago, Chile*
[2] *Institute of Cell, Animal and Population Biology, University of Edinburgh Ashworth Laboratories. King's Buildings, Edinburgh, UK*
[3] *School of Biological Sciences, Washington State University, USA*
[4] *Rolf Nevanlinna Institute, Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland*

(*Received 12 July 2004 and in revised form 25 May and 23 September 2005*)

Summary

A Bayesian model and variable dimensional parameter estimation based on Markov chain Monte Carlo was applied to map quantitative trait loci (QTLs) in a doubled haploid mapping population of rainbow trout. To increase power, the analysis was performed using the multiple-QTL model, which simultaneously accounted for all the environmental and genetic main effects that influence the expression of early development life history traits. By doing so we obtained the posterior estimated effects for the environmental factors as well as the number, positions, and the effects for the QTLs. The analyses revealed QTLs for time at hatching, embryonic length and weight at swim-up stage. The posterior expectation of the number of QTLs in different linkage groups shows that at least four QTLs are needed to explain the observed differences in early development between the clonal lines. The Bayesian method effectively combined all the information available to accurately position these QTLs in the rainbow trout genome.

1. Introduction

During the last decade, quantitative trait loci (QTLs) have been mapped using linkage disequilibrium generated by crosses between inbred (or outbred) lines that differ widely in their values for the trait of interest. In classical QTL analysis using interval mapping, the LOD-score profile is constructed over different genomic positions and the highest LOD value in each chromosome (which is also higher than a predetermined significance threshold) is taken as evidence of putative QTL position. The estimate of the position of the QTL obtained in this way generally coincides with the maximum likelihood estimate of position (Sorensen & Gianola, 2002). Alternatively, the full maximum likelihood (ML) analysis of Lander & Botstein (1989) can be approximated by regressing QTL genotype probabilities onto the phenotypes and using least squares (LS) estimation available in most statistical packages (Martinez & Curnow, 1992; Haley & Knott, 1992). The genetic model used in interval mapping can be expanded to include simultaneous contributions of more than one QTL. Inferring the number of QTLs and estimating their genomic positions requires model selection and search strategies (Sillanpää & Corander, 2002; Broman & Speed, 2002). There are also methods which approximate QTLs by fitting markers or 'virtual markers' as covariates in order to capture background variation caused by QTLs other than the one tested (Zeng, 1994; Jansen, 1993).

Bayesian analysis provides a straightforward framework in which the posterior distributions of all the unknowns (parameters and missing values) in the QTL mapping problem can be estimated using Markov chain Monte Carlo (McMC) sampling methods. Under this framework, rather than maximizing the likelihood of obtaining the maximum likelihood (point) estimate of the parameter, inference

* Corresponding author. Tel: +56 2 978 5597. Fax: +56 2 978 5611. e-mail: vmartine@uchile.cl

† Present address: Department of Biological Sciences and Center for Reproductive Biology University of Idaho, USA.

is based on the whole (posterior) distribution of parameters (Hoeschele *et al.*, 1997; Hoeschele, 2001; Sorensen & Gianola, 2002; Tanner, 1996). The posterior distribution is obtained by combining prior distribution with the likelihood of the data (the sampling model) through Bayes' formula (Shoemaker *et al.*, 1999). The analytical solution of the posterior distribution necessitates integrating (or summing) over high-dimensional parameter spaces, which is often impossible in practice. These integrals can be approximated using McMC. In particular, the Bayesian model implemented via McMC provides a means of treating the number of parameters of a model as an unknown, to be inferred from the data at hand (Green, 1995).

Surprisingly, in many QTL mapping analyses environmental (non-genetic) effects are not included explicitly in the genetic model used for analysis (see, for example, experiments with various objectives, e.g. De Koning *et al.*, 1999; Bidanel *et al.*, 2001; Hawthorne & Via, 2001). One usual practice is to first adjust phenotypic observations for environmental factors and then carry out a separate QTL analysis where residuals of the first analysis are treated as phenotypes, either with (residual) maximum likelihood or least squares (Basten *et al.*, 2002). By doing so, uncertainty in estimates of environmental factors is underestimated and power of subsequent QTL analysis could be reduced. This is especially true when the number of environmental factors is large. On the other hand, the simulations of Martinez (2003) suggest that if environmental factors are omitted completely from the model, the power loss is clear compared with the joint analysis, whose power, on average, is constant irrespective of the actual magnitude of the environmental effects.

The objective of the present study was to obtain the posterior distribution of all the parameters of interest in the QTL mapping problem, such as a number, location and effects of QTLs, as well as of the environmental factors in a doubled haploid population of rainbow trout. We adopted a Bayesian mapping approach, designed for inbred line crosses, to obtain posterior samples of all the unknowns given the data and the prior distribution (Sillanpää & Arjas, 1998). We incorporated environmental covariates into the genetic model of Sillanpää & Arjas (1998).

The benefit of the joint analysis approach in terms of accuracy of position has been shown previously using simulation (Martinez, 2003). In this paper we present a real-data analysis of a mapping population derived from crosses between clonal lines of rainbow trout that differ in the rate of early development. The data have been analysed previously without incorporating environmental cofactors (Robison *et al.*, 2001; Martinez *et al.*, 2002*b*).

2. Materials and methods

(i) *Implementing the Bayesian method for the doubled haploid design with environmental effects*

(a) *Genetic model*

Let us consider the analysis of the doubled haploid design as obtained from a cross between clonal lines. We use the extended model of Sillanpää & Arjas (1998), where we include control of environmental conditions and covariates (e.g. age, sex or treatment). It is assumed that the trait, conditionally on effects of QTL and environment, follows a normal distribution (i.e. this is used for construction of the likelihood). The marker map is assumed to be known and some fraction of marker genotypes may be missing. The notation used throughout is: vector of phenotypes for the quantitative trait (y), the number of QTL (N_{QTL}), the number of genotypes ($N_{GEN}=2$, $a_1=$AA and $a_2=$BB, for the two homozygous genotypes of the doubled haploid design), QTL genotype matrix (X), i.e. the QTL genotype of all individuals obtained from the current location sampled, genotypic effects of QTL (b_{qj} for the jth genotype of QTL q) and of background controls (BC; c_{kj} for the jth genotype of background control k, which are included in the model to control variation of QTLs in other linkage groups (Zeng, 1994; Jansen, 1993; Sillanpää & Arjas, 1998). In this setting, the following over-parameterized regression model is used to explain the observation for individual i:

$$y_i=\rho' B_i+\sum_{q=1}^{N_{QTL}}\sum_{j=1}^{N_{GEN}} b_{qj}1_{\{x_{qi=\alpha_j}\}}$$

$$+\sum_{k=1}^{N_{BC}}\sum_{j=1}^{N_{GEN}} c_{kj}1_{\{X^*_{ik}=\alpha_j\}}+e_i.$$

Here, $1_{\{x_{qi}=\alpha_j\}}$ and $1_{\{X^*_{ik}=\alpha_j\}}$ are indicator variables, such that they take values of one when the individual i is of genotype α_j and zero otherwise (see Sillanpää & Arjas, 1998). We assume that $\rho=(\rho^{(1)}, \rho^{(2)})$ (environmental effects) is a vector of regression coefficients of quantitative covariates $\rho^{(1)}$ and class means of qualitative covariates $\rho^{(2)}$. Moreover, $B_i=(B_{ci})$ is the vector of environmental covariates for individual i. In case of no covariates, the model reduces to the original (including the intercept). More specifically,

$$\rho' B_i=\sum_{c=1}^{N_L}\rho_c^{(1)}B_{ci}+\sum_{c=N_L+1}^{N_L+N_C}\sum_{j=1}^{N(c)}\rho_{cj}^{(2)}1_{\{B_{ci}=j\}}.$$

Here, N_{L} is a number of linear regression (quantitative) covariates and N_C is a number of classification (qualitative) covariates. Similarly, a vector of regression coefficients ρ is divided into linear regression coefficients $\rho^{(1)}=(\rho^{(1);\,c})$, $\{c=1, \ldots, N_L\}$ and class means $\rho^{(2)}=(\rho^{(2);\,cj})$ $\{c=N_L+1, \ldots, N_L+N_C,$

https://doi.org/10.1017/S0016672305007871 Published online by Cambridge University Press

$j = 1, \ldots, N(c)\}$, where $N(c)$ is a number of categories in covariate c. Indicator variable $1_{\{B_{ci}=j\}}$ takes the value one when the value of covariate c in individual i belongs to category j, and zero otherwise.

(b) *Priors*

Independent bounded uniform prior distributions for most of the parameters (including the environmental covariates) are used in order to diminish as much as possible the influence of the priors on the posterior. For the number of QTLs, the prior distribution used was an accelerated truncated Poisson with mean equal to 0·76 (Gaffney, 2001; Jannink & Fernando, 2004; Sillanpää *et al.*, 2004). The QTL genotype probabilities (prior) were calculated conditionally on flanking markers using an algorithm for doubled haploid lines which is actually the same as that used previously for the backcross design (Knapp *et al.*, 1990; Sillanpää & Arjas, 1998).

(c) *Estimation*

As in Sillanpää & Arjas (1998) the Bayesian model is fitted using McMC sampling to estimate the parameters. In this paper, instead of sampling each parameter at a time, the regression parameters are here updated in two separate blocks, which are described below in detail, as (1) and (2). In general, use of block-updating for a set of parameters can improve the mixing properties of the chain when highly correlated variables are sampled (Sorensen & Gianola, 2002; Richardson & Spiegelhalter, 1996). The reversible jump McMC is used to make estimation and model selection simultaneously. During the estimation, the algorithm visits (or jumps) between model dimensions, corresponding to different numbers of QTLs, and the time spent in each model (on average) is proportional to the posterior probability of the model. In other words, this strategy of jumping between model dimensions (during the simulation) allows one to collect samples to summarize the relative importance of each model. This algorithm is an extension of the Metropolis–Hastings algorithm, where the acceptance probabilities are of the form *min*[1, (posterior ratio) × (proposal ratio) × (Jacobian of the transformation)]. A more detailed explanation can be found elsewhere (Hoeschele, 2001; Waagepetersen & Sorensen, 2001; Gaffney, 2001). Here we summarize the changes to the sampling strategy of Sillanpää & Arjas (1998).

(1) The intercept, residual variance and environmental effect coefficients form a single block. The random walk (Chib & Greenberg, 1995; Richardson & Spiegelhalter, 1996) proposal is generated for each parameter by using a uniform symmetric distribution around the current value ($x \pm \varphi$). The value of φ is chosen such that the acceptance rate enables sufficient mixing of the chain (Chib & Greenberg, 1995; Richardson & Spiegelhalter, 1996) in order to achieve convergence to the stationary (posterior) distribution.

(2) All the QTLs and background control effect coefficients are updated together as a single block. Their proposals are also generated by using random walk.

If a covariate is of a categorical type then its first effect coefficient is restricted to zero, and the value zero is repeatedly proposed for the corresponding coefficient in each round. The acceptance of the block proposal is verified by using the ratio of likelihoods evaluated at the new and current values, respectively. If a proposal is not accepted, the whole block is discarded and current values are maintained.

The posterior distribution of the QTL effects was estimated only for those regions that showed a high QTL intensity, conditional on the interval of 1 cM around the peak of the QTL intensity, i.e. in the more informative areas of the linkage groups (Fig. 4). Note, however, that in practice the posterior distributions here differ little whether or not all the information from the linkage group was used or not. Due to the fact that we used an over-parameterized regression model (as explained above), the point estimate of the posterior distribution of additive QTL effects was calculated as the mean of the difference between the homozygous effects from the original clonal lines used to generate the mapping population.

At every iteration, a 'new' QTL is accepted (or rejected/modified) when applying the reversible jump McMC algorithm and therefore we do not keep track of any 'particular' QTL. This means that the location of a particular QTL is irrelevant. Instead, the location of the QTL in the chromosome is obtained as the proportion of times in which one (or more) QTL is assigned to a particular interval: this is the so-called QTL intensity. This is not a posterior distribution of locations of the QTL; rather it is the posterior probability that a QTL is located in each particular interval (of a certain length), which is in fact a model-averaged estimate of location in the linkage group (across the number of QTLs). See Hoti *et al.* (2002) for an alternative formulation.

(ii) *Doubled haploid mapping population of rainbow trout*

The doubled haploid mapping population was developed at Washington State University (Pullman) from an all-male (XY) F_1 population obtained using crosses between two clonal lines, OSU and Swanson (SW), that differed in the average time at hatching. The clonal lines were formed by two successive generations of androgenesis. In the first generation, parents were sampled from the outbred population

https://doi.org/10.1017/S0016672305007871 Published online by Cambridge University Press

in order to produce homozygous, individually distinct doubled haploid progeny, from which males or females were selected to form the clonal lines. The expected sex ratio of androgenetic progeny is equal to 50% for males and females, since males are heterogametic. A second generation of androgenesis or gynogenesis was used to produce the clonal lines that were raised in the laboratory. From this set of clones, two lines were selected in order to produce the F_1 population (Robison *et al.*, 1999). The first is an all-male line derived from a natural population from Alaska (YY-SW) and the second is an all female (XX-OSU) line obtained from a domesticated strain obtained from Oregon State University (Robison *et al.*, 1999). These clonal lines show divergent hatching times in the laboratory, with the SW line showing accelerated early development. The isogenicity of these clonal lines has been confirmed using DNA fingerprinting (Robison *et al.*, 1999). The sperm of these F_1 individuals was used to obtain two groups of androgenetic rainbow trout (group 1, $n=57$; group 2, $n=149$) that were raised at slightly different temperatures (10·4 °C versus 11·6 °). Successful androgenesis from a F_1 parent resulted in diploid organisms that contained two sets of identical paternal chromosomes with an equal proportion of male (YY) and female (XX) individuals. Three traits related to early development were examined: time at hatching (TTH: mean 705·6 h, SD = 76·8 h), embryonic length (LEN: mean 63·0 mm, SD = 19·8 mm) and weight (WEI: mean 20·2 mg, SD = 2·5 mg). The marker map comprised a total of 27 linkage groups, spanning 974·6 cM (about 40% of the rainbow trout genome), obtained from segregation of 222 AFLP markers. The mean distance between markers in the linkage groups was about 8 cM, with 10 cM standard deviation. A more detailed outline of the procedure to produce the AFLP marker map is presented by Robison *et al.* (2001), and power-related issues regarding doubled haploid individuals were discussed by Martinez *et al.* (2002*a*).

Table 1. *Prior uniform distributions used in the analysis of time of hatching* (*TTH*), *embryonic length* (*LEN*) *and weight* (*WEI*)

	Trait		
Parameter	TTH	LEN	WEI
Intercept	$U_{(-500,\ 500)}$	$U_{(-10,\ 10)}$	$U_{(-50,\ 50)}$
s^2	$U_{(0,\ 2000)}$	$U_{(0,\ 6\cdot5)}$	$U_{(0,\ 387)}$
QTL coefficients	$U_{(-100,\ 100)}$	$U_{(-10,\ 10)}$	$U_{(-20,\ 20)}$
BG coefficients	$U_{(-100,\ 100)}$	$U_{(-10,\ 10)}$	$U_{(-20,\ 20)}$
Location	$U_{(0,\ 120)}$	$U_{(0\cdot0,\ 120)}$	$U_{(0\cdot0,\ 120)}$

An initial analysis, performed using the forward and backward stepwise regression analysis as implemented in QTL Cartographer (Basten *et al.*, 2002), revealed six linkage groups showing association between markers and phenotypes. These groups were selected for further analysis using the Bayesian method, and these markers were used as cofactors to control background genetic variation in linkage groups other than the one currently analysed. This greatly reduces the computational demand of the method (but see Section 4). During incubation, development may be disrupted giving rise to the presence of mild deformities. A preliminary analysis showed that including (and using the presence of deformities as a fixed effect in the analysis) or excluding these individuals in the analysis of length gave very similar results in terms of position of the QTL in the linkage groups pre-selected, but the likelihood ratio test scores at the most likely position were much lower when individuals were discarded. The slight increase in environmental temperature between the groups also produced a significant effect in TTH. For these reasons, in the final analysis the presence of deformities and temperature at incubation were used as an environmental covariate for LEN and TTH, respectively. The range of proposal distributions was specified after several preliminary test runs in each of the linkage groups analysed. For the final analysis, proposals giving adequate mixing of the chains were used to run a single long chain (1×10^6), All samples were retained for further analysis. Features of these posterior distributions can be analysed using standard measures of convergence such as that of Geweke (1992). This analysis relies upon the fact that if the chain is stationary then the means of the first and last part of the chains should be similar. The Z statistic is calculated as the difference between the two means divided by the asymptotic standard error, where the variance is obtained in such a way that the correlation between samples is accounted for, using spectral density estimation (Geweke, 1992; Tanner, 1996; Smith, 2003). The ranges of the prior distributions for the different traits are presented in Table 1.

3. Results

(i) *Mixing properties of the chain*

For different traits and chromosomes, we performed a series of exploratory analyses in order to find suitable ranges of the proposal distributions. This is paramount for obtaining adequate mixing properties of the sampler. The mixing was monitored by visually inspecting the sample paths of different parameters and using the cumulative occupancy posterior probabilities of a model with 0, 1, 2 or 3 QTLs to monitor convergence of the number of QTLs (Heath, 1997; Uimari & Sillanpää, 2001). After this tuning stage,

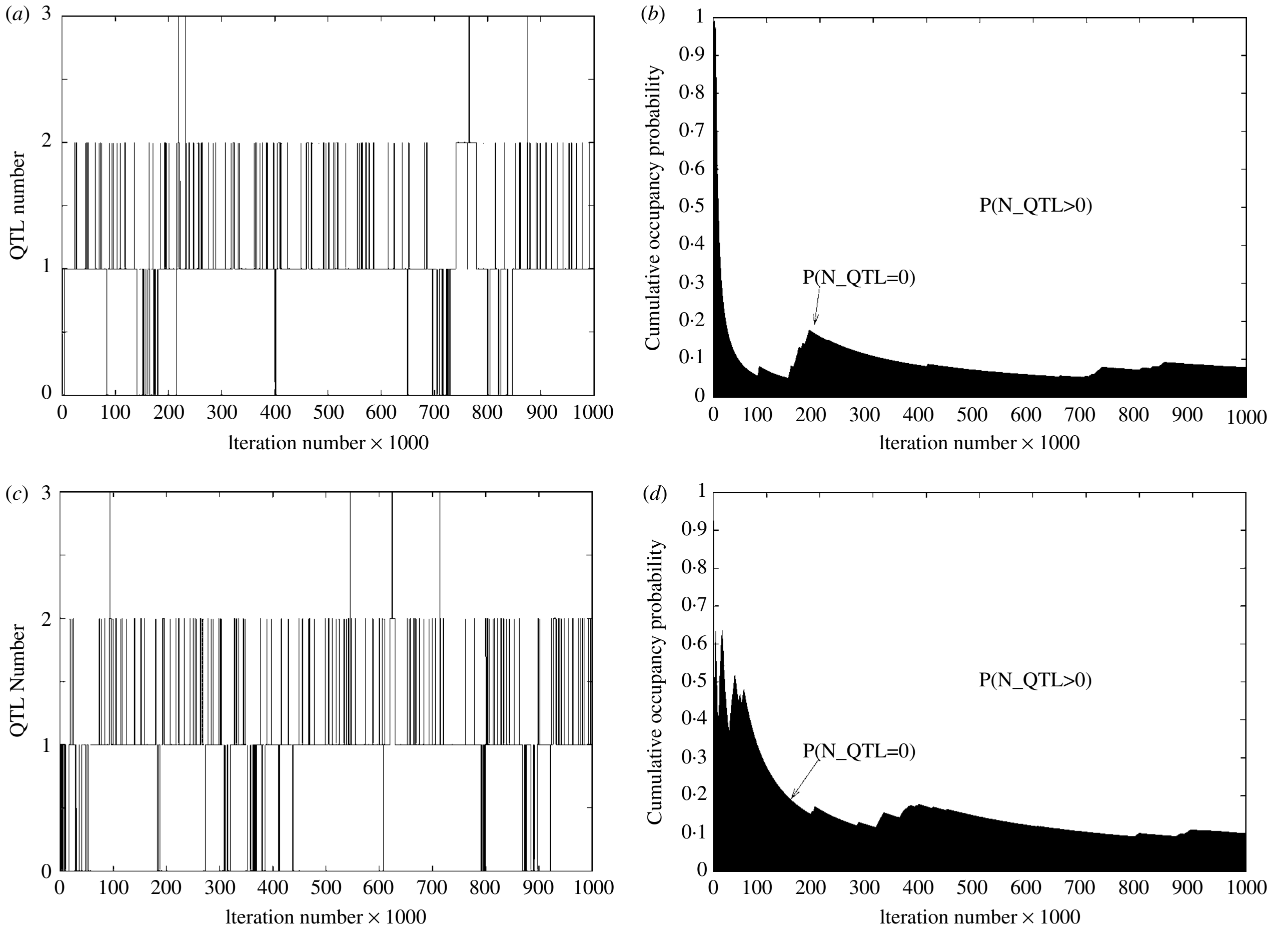

Fig. 1. Sample paths (*a*, *c*) and cumulative occupancy probabilities for different QTL models (*b*, *d*, according to the number of QTLs) obtained from linkage group V for time at hatching (TTH) (*a*, *b*) and embryonic length (LEN) (*c*, *d*).

the McMC sampler was able to move well across the entire parameter space of the number of QTLs. Fig. 1*a* and *c* present an example of the sample paths for TTH and LEN on linkage group V. Changes between a single-QTL and a two-QTL model were more regular than changes from the models of zero or three QTLs. Based on the cumulative occupancy probabilities, it was apparent that, after an initial period where the cumulative probability of a zero-QTL model was as likely as the probability of a single-QTL model, this state moves rapidly towards a distribution in which a single QTL is many times more likely than a model in which there is no QTL (Fig. 1*b*). A similar degree of mixing was found for other traits and linkage groups.

The same behaviour was seen for other parameters in the model. It was clear that after the tuning stage the values of the proposal ranges enable enough mixing to obtain posterior distributions that converge to a stationary distribution. The value of the proposal for the QTL effect was the most important to tune in order to obtain adequate mixing of the chains. Acceptance rates for most of the parameters were between 0·23 and 0·50. The trace plot shows that the chain moves quickly into the region where the posterior mass was found for the QTL effect in linkage group V (Fig. 2). The Geweke convergence diagnostic shows that there is no difference in the mean between the first 10% and the last 50% of the observations of the chain with a Z value of 1·4 ($P=0{\cdot}16$). The same was found for time at hatching in the same linkage group ($Z=0{\cdot}91$; $P=0{\cdot}36$).

(ii) *Posterior distribution of environmental parameters*

In general, features of the distribution of the contrast of the two levels of the environmental effects were quite similar for all the linkage groups, irrespective of whether there was evidence of linkage or not. The posterior distributions overlap substantially, as shown in Table 2. For TTH the slight change in temperature between the two groups decreased the time of hatching greatly, with a posterior mean for the contrast over all linkage groups equal to 94·4 h. As expected, the presence of deformities decreased the length of the fry; the estimated posterior mean of the contrast between the two levels of these effects

https://doi.org/10.1017/S0016672305007871 Published online by Cambridge University Press

Table 2. *Posterior estimates of the mean, variance* (s^2) *and 95% credible regions* (*95% CR*) *for the environmental regression coefficients fitted* (b_e) *and residual variance* ($s^{2;e}$) *of time of hatching* (*TTH*) *and embryonic length* (*LEN*) *and weight* (*WEI*)

Trait		Mean (b_e)	s^2 (b_e)	95% CR (b_e)	Mean ($s^{2;e}$)	s^2 ($s^{2;e}$)	95% CR ($s^{2;e}$)
TTH	LG-IV	−90	39	[−105; −77]	1294	25 790	[973; 1667]
	LG-V	−94	38	[−107; −79]	1226	31 840	[892; 1629]
	LG-XII	−95	36	[−108; −81]	1169	23 849	[862; 1503]
	LG-XV	−99	56	[−114; −81]	1001	37 662	[836; 1434]
LEN	LG-V	−2·72	0·09	[−3·34; −2·21]	2·79	0·11	[2·14; 3·56]
	LG-XI	−2·62	0·09	[−3·28; −2·02]	3·06	0·10	[2·48; 3·78]
	LG-XII	−2·64	0·09	[−3·32; −2·01]	2·89	0·13	[2·18; 3·64]
			I				
WEI	LG-VI	19·7	7·5	[14·1; 25·9]	256	988	[200; 328]
	LG-X	20·0	6·7	[14·8; 25·6]	251	784	[199; 314]
	LG-XV	20·4	6·2	[15·3; 25·7]	229	713	[183; 293]

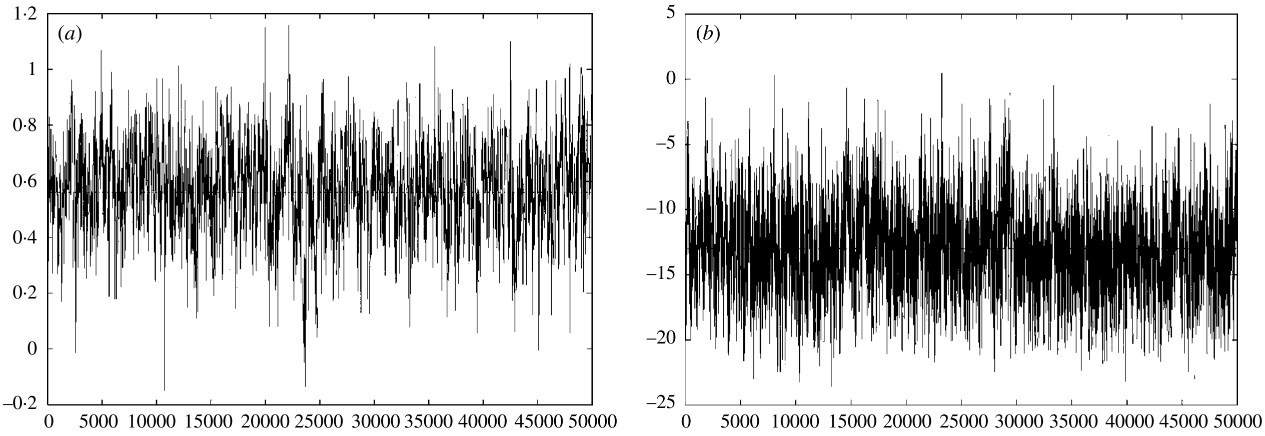

Fig. 2. Description of the McMC sampler (trace plots) of QTL effects for time at hatching (TTH) and embryonic length (LEN) in linkage group V. (*a*) QTL effect LEN; LG-V. (*b*) QTL effect TTH; LG-V.

was about −2·65 mm. There was only a slight increase in the posterior mean of the residual standard deviation in cases where the linkage group did not show evidence of linkage and a small decrease when there was evidence of linkage.

(iii) *Posterior probability of model parameters and location of putative QTLs*

The posterior probabilities of number of QTLs in different linkage groups are presented in Table 3 for the three traits analysed. The posterior distribution of number of QTLs can be interpreted directly as the posterior probability of linkage for the model in which at least a single QTL or zero QTL (no linkage) is present. Although these posterior distributions summarize all the information regarding QTL activity in the linkage groups, this does not necessarily mean that the QTLs can be positioned accurately in the linkage groups. For this reason, the interpretation of the whole Bayesian analysis requires that the marginal posterior distribution of all the parameters in the model be assessed jointly.

(a) *Posterior probability of the number of QTLs*

The analyses of the traits LEN and TTH both strongly supported presence of a single QTL in linkage groups V and XII. In other words, these linkage groups had large probabilities for a single QTL and small probabilities for no QTL. In the trait WEI, the evidence for a single QTL was much weaker, except on linkage groups X and XV which both provided a clear signal. Additionally in the trait TTH and linkage group XV, the posterior probability of two QTLs is 5 (or even higher) times more likely than 0, 1 or 3 QTLs. Two QTLs also appeared highly probable for the trait WEI in linkage group VI. As we will show below, these QTLs appear to be isolated, without having any evidence of QTL activity in adjacent

https://doi.org/10.1017/S0016672305007871 Published online by Cambridge University Press

Table 3. *Posterior probability of number of QTL* (N_{QTL}) *given the data in the different linkage groups* (*LG*)

			LG				
Trait	No. of QTL	Prior	**V**	**IX**	**XI**	**XII**	**XV**
LEN	0	0·402	0·101	0·802	0·862	0·010	0·873
	1	0·454	0·869	0·192	0·134	0·912	0·119
	2	0·128	0·031	0·006	0·004	0·078	0·008
	3	0·016	0·000	0·000	0·000	0·000	0·000
	E(N_{QTL}|data)	0·758	0·93	0·20	0·14	1·07	0·13
TTH	0	0·402	0·025	0·205	0·568	0·000	0·000
	1	0·454	0·905	0·656	0·242	0·968	0·159
	2	0·128	0·065	0·137	0·173	0·032	0·809
	3	0·016	0·005	0·002	0·018	0·000	0·031
	E(N_{QTL}|data)	0·758	1·05	0·94	0·64	1·03	1·87
WEI	0	0·402	0·521	0·085	0·277	0·649	0·352
	1	0·454	0·301	0·095	0·640	0·222	0·513
	2	0·128	0·148	0·713	0·077	0·127	0·127
	3	0·016	0·030	0·107	0·005	0·009	0·009
	E(N_{QTL}|data)	0·758	0·69	1·84	0·81	0·50	0·79

intervals surrounded by the two QTL (Whittaker *et al*., 1996).

The posterior probabilities of no QTL for TTH were relatively high in linkage group XI. In linkage group IX, the probability of a single QTL is equal to 0·65, which is 3 times the value for no QTL (Table 3). However, this result should be interpreted with caution since the QTL intensity was bimodal with two peaks at marker positions 82 and 86 (results not shown). In addition, 95% credible regions overlap was 0 (mean (SD), −7·4 (8·5); median, −7·5; 95% CR, −25·44, 11·17). The most parsimonious interpretation of these results is that there is not sufficient information in the data to accurately summarize the underlying process or there are other sources of genetic variation present in the data (i.e. interaction between QTLs). This is consistent with the fact that the mean estimate of the posterior distribution of the QTL effect is only a third of the mean value obtained in other linkage groups, suggesting that this QTL has a small effect. A larger sample size would be needed in order for the data to be informative enough to confirm this preliminary finding.

(b) *Location estimates from QTL intensity*

The QTLs were located using the posterior QTL intensity for those linkage groups that showed a high posterior probability of at least one QTL segregating (Fig. 3). Note that there is a broad area in which the QTL may reside for TTH in LG-V. This may be due to the fact that the marker bracket in which the QTL was located is rather large (about 50 cM) so that there may not be enough information to precisely locate this QTL. The most likely position is at 107·8 cM between the marker positioned at 67·8 cM and the right telomeric marker of this linkage group. In the same linkage group, the most likely position of the QTL for LEN was found at position 20·8 cM, almost in the middle of the second interval; about 53% of all the hits were found in this interval. About 33% of all the iterations in which there was at least a single QTL sampled fall in the same interval as found for TTH. This suggests that there may be two linked QTLs influencing LEN and TTH in this linkage group, and a second with pleiotropic effects on TTH and LEN at the right end of the chromosome. However, no evidence of more than a single QTL for LEN and TTH was obtained in the present analysis (see Table 3).

The most likely positions for the QTLs for LEN and TTH on linkage group XII were very similar: 14·9 and 10·7 cM, respectively (Fig. 3). The QTL intensities overlap substantially, even though the QTL intensity for LEN was much more peaked. This suggests that the effects of these two QTLs could arise from a single pleiotropic locus (see Fig. 3). Further support for this interpretation is provided by Martinez *et al*. (2002*b*).

On linkage group XV the posterior probability of the number of QTLs was sharply centred at two QTLs. This finding is consistent with what is obtained for the QTL intensity in this linkage group. Two modes were found at marker positions 36·7 and 75·5 cM on linkage group XV. Note that the posterior QTL intensities were very peaked, presumably due to the fact that at marker positions there is more information to detect a QTL when it is completely linked with the marker. Note that the QTL intensity is multi-modal in the region between 23 and 45 cM. In general, some discontinuity in the posterior QTL intensity can be observed at marker positions when

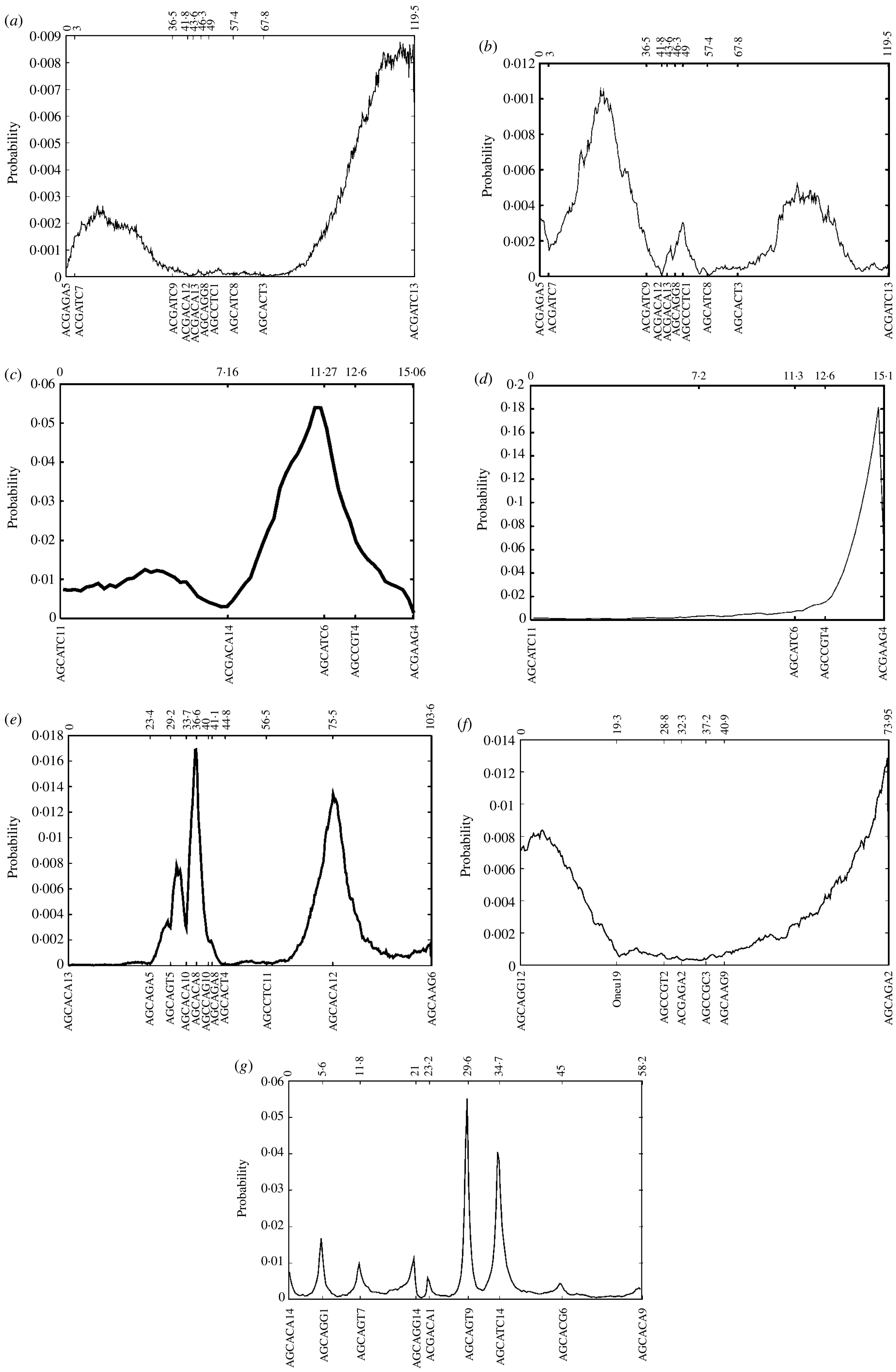

Fig. 3. For legend see opposite page.

https://doi.org/10.1017/S0016672305007871 Published online by Cambridge University Press

the QTL is not completely linked with the marker, as has been shown previously (Bink *et al*., 2000; Hoti *et al*., 2002). Note that when using the Bayesian method there is not the bias expected towards markers as found when using the bootstrapping method in cases where the power for detecting the QTL is low (Yi & Xu, 1999). In our analysis, the posterior QTL intensity falls sharply at the position where the next marker to the left is situated in this linkage group.

Although on linkage group X there is evidence of a single QTL (the probability of a single QTL doubled the probability of no QTL), the location was quite inaccurate since bimodality of the QTL intensity was seen in the interval between the sixth and the seventh markers in this linkage group (Fig. 3).

(c) *Posterior distributions of QTL effects*

In general, there was sharp posterior information for QTL effects, and the distributions were peaked and quite symmetrically distributed around the mean (Fig. 4).

The putative QTL detected for TTH explained in total about 40% of the observed variance and these QTLs explained most of the difference between the two clonal lines, which was about 2·5 standard deviations, calculated from results of Robison *et al.* (1999). In linkage group XV, the point estimates of the posterior distribution of QTL effects have opposite effects, decreasing the hatching time (as calculated from the first mode of the QTL intensity at position 36·7 cM) and the other of similar magnitude but with positive sign. It is likely that the QTLs were in repulsion in the F_1 population.

About 18% of the total variance is explained by the effects as obtained from linkage group V and XII for LEN. The point estimates have the same sign and explain a relatively similar percentage of the observed variance. The information obtained from WEI in linkage group VI and X explains overall around 20% of the phenotypic variance. The two telomeric QTLs in linkage VI have opposite signs, suggesting that both were in repulsion in the F_1.

4. Discussion

The objective of the present study was to illustrate the utility of joint analysis of environmental and genetic effects in QTL mapping and to obtain posterior evidence of segregation of QTL in a doubled haploid mapping population. We prepared a new version of the Bayesian model of Sillanpää & Arjas and incorporated all the environmental data available in the experiment in order to obtain the posterior distribution of the number of QTL, environmental effects, locations and effects of the different QTL that are segregating. The data were derived from a cross between clonal lines that diverge in early development rate that has previously been analysed without including environmental effects explicitly in the model (Robison *et al*., 2001; Martinez *et al*., 2002*b*). By including the environmental effects in the Bayesian model we obtained further evidence of other multiple QTL segregating for the different traits examined. From an evolutionary perspective the study of early development in natural populations is important, since these traits can have fitness implications through their effect on timing emergence from the redd, which is a milestone of the population dynamics of many salmonid species (Robison *et al*., 1999). Previous work used only phenotypes to study the genetics of these traits. This study showed evidence of major genes segregating which explain a large proportion of the observed differences between the clonal lines in development rate.

The Bayesian analysis used cofactors selected using a forward and backward elimination (stepwise regression) procedure implemented in QTL Cartographer (Basten *et al*., 2002). Such pre-selection was done for computational efficiency. In general, however, the use of marker cofactors may be not ideal since, especially in small populations, unlinked markers can be correlated with markers in the chromosome under investigation. Fitting markers in other chromosomes has a high impact on the posterior probability of linkage of the current chromosome, i.e. fitting a cofactor can change the evidence of linkage in contrast to when no marker cofactors are included (Maliepaard *et al*., 2001). A sensible alternative would be to model the genome in a single multiple QTL analysis and, rather than include marker cofactors as 'known' quantities, to include them as 'unknowns', further reversible jumps being required to add or delete QTLs in other chromosomes. Further investigation is required to test this alternative. Nevertheless, this analysis would be computationally very expensive and the convergence would be complicated to assess, which may impair its applicability in practical situations. More suitable alternatives in this respect are provided by the recent Bayesian shrinkage estimation methods (Xu, 2003; Wang *et al*., 2005).

We have used a simple genetic model for QTL detection in which only additive effects are considered.

Fig. 3. Posterior QTL intensities on linkage groups showing a high posterior probability of at least one QTL for the different traits analysed. (*a*) TTH; LG-V. (*b*) LEN; LG-V. (*c*) TTH; LG-XII. (*d*) LEN; LG-XII. (*e*) TTH; LG-XV. (*f*) WEI; LG-VI. (*g*) WEI; LG-X.

https://doi.org/10.1017/S0016672305007871 Published online by Cambridge University Press

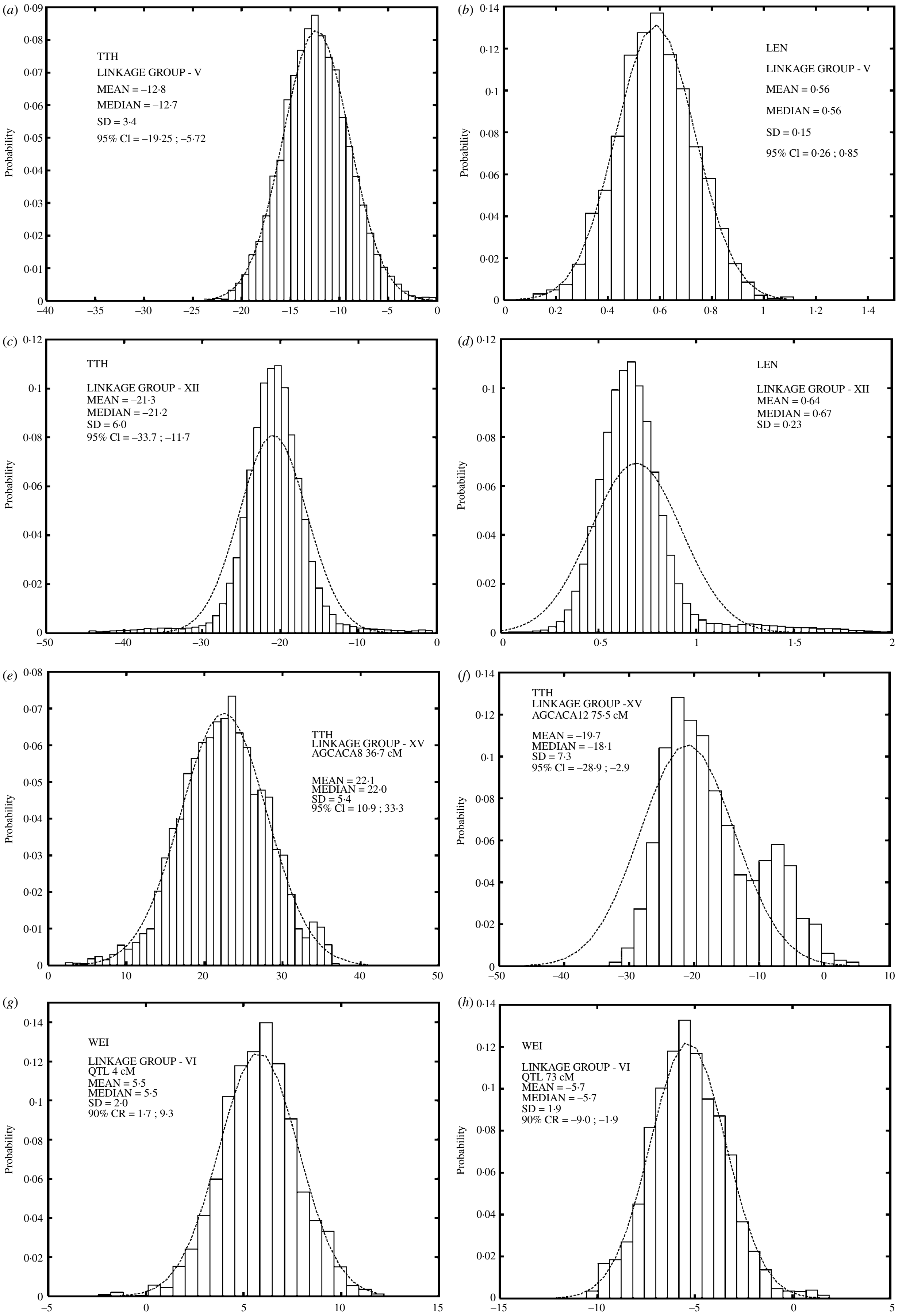

Fig. 4. For legend see opposite page.

https://doi.org/10.1017/S0016672305007871 Published online by Cambridge University Press

More complex models such as those involving interaction between QTL can also be implemented. However, this would require analysing the genome as a whole in order to allow interactions between QTLs in different linkage groups. Recently, Yi & Xu (2002) developed a Bayesian model in which pairwise interactions within a single linkage group are always sampled for the epistatic model given that at least two QTLs (with marginal effects) are accepted in the linkage group and inferences about QTL effects are carried out by inspecting the posterior distributions (see also Conti *et al.*, 2003). An optimal strategy that deserves further investigation would be to include the epistatic effect as a variable in the model and apply a reversible jump McMC algorithm (Yi *et al.*, 2003; Narita & Sasaki, 2004). Again, an interesting alternative is provided by the recent Bayesian shrinkage estimation methods (see Zhang & Xu, 2005), but also application of partition approaches (see Seaman *et al.*, 2002) deserves further consideration.

We have used AFLP markers in order to provide a large number of marker for the QTL analysis (Vos *et al.*, 1995; Young *et al.*, 1998). The main advantage of using the AFLP technique is that it does not require previous knowledge of the DNA sequence, generates fingerprinting profiles that can be reproduced and allows the amplification of a high number of DNA fragments per reaction, thus enabling the cost-effective detection of specific amplified fragments (Alves *et al.*, 2002). However, because they can only be scored as presence or absence, their utilization in other types of designs, such as F_2 or four-way crosses or outbred populations requires modification of the algorithms used for QTL mapping analysis (Gessler & Xu, 1999).

The use of doubled haploids for detecting QTLs, provided that clonal lines are available, has been shown to be much more efficient than a standard F_2 design (Martinez, 2003). The increase in power is due to the large increase in the genetic variance in the doubled haploid population, which is double that expected in the F_2 because the genotype frequencies in the doubled haploid population are redistributed compared with the F_2. One disadvantage, however, is that it is not possible to estimate other important sources of genetic variation such as dominance. Another potential complication is the possibility of segregation distortion, which is most likely due to association between markers and deleterious mutations. In our experiment, severely deleterious mutations were likely to be eliminated when forming the clonal lines. We found no evidence of segregation distortion in the linkage groups under analysis (data not shown). This type of design can be used to map viability genes (Ritland, 1996; Vogl & Xu, 2000; Luo & Xu, 2003) and it is expected that this analysis would have more power than classical designs used to study deleterious mutations (McCune *et al.*, 2002). In the long term, the clonal populations produced through androgenesis may provide a unique source for studies on accumulation of deleterious mutations in the absence of recombination (Guex *et al.*, 2002). A related approach has been successfully used for mapping viability genes in the tilapia (Palti *et al.*, 2002).

A new version of the Multimapper software (version 1.0), used for analyses in this paper, is now generally available for research purposes from http://www.rni.helsinki.fi/∼mjs.

We are grateful to Dr Sara Knott, Professor Bill Hill, Professor Brian Charlesworth and two anonymous referees for comments that greatly improved and clarified the manuscript. We acknowledge support from the University of Chile (DID) hatch project 03/03-2, the Academy of Finland grant no. 202324 (M.S.), the National Institutes of Health (NIEHSPO1ES04766 and BMMRP1RO1-RR06654) and the National Science Foundation (award IBN 0082773).

References

Alves, E., Castellanos, C., Ovilo, C., Silio, L. & Rodriguez, C. (2002). Differentiation of the raw material of the Iberian pig meat industry based on the use of amplified fragment length polymorphism. *Meat Science* **61**, 157–162.

Basten, C., Weir, B. & Zeng, Z. (2002). QTL Cartographer, version 1.16. Raleigh, NC: Department of Statistics, North Carolina State University.

Bidanel, J., Milan, D., Iannuccelli, N., Boscher, Y., Bourgeois, M., Caritez, F., *et al.* (2001). Detection of quantitative trait loci for growth and fatness in pigs. *Genetics, Selection and Evolution* **33**, 289–309.

Bink, M., Janss, L. & Quaas, R. (2000). Markov chain Monte Carlo for mapping quantitative trait loci in outbred populations. *Genetical Research* **75**, 231–241.

Broman, K. & Speed, T. (2002). A model selection approach for identification of quantitative trait loci in experimental crosses (with discussion). *Journal of the Royal Statistical Society, Series B* **64**, 641–656.

Chib, S. & Greenberg, A. (1995). Understanding the Metropolis–Hastings algorithm. *American Statistician* **49**, 327–335.

Conti, D., Cortessis, V., Molitor, J. & Thomas, D. (2003). Bayesian modeling of complex metabolic pathways. *Human Heredity* **56**, 83–93.

De Koning, D., Janss, L., Rattink, A., Oers, P. V., Vries, B. D., Groenen, M., *et al.* (1999). Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (*Sus scrofa*). *Genetics* **152**, 1679–1690.

Fig. 4. (*a*)–(*h*) Posterior distribution of QTL effects conditional on the centimorgan where the mode of the posterior QTL intensity was observed. The smooth curves were obtained from the normal distribution using the posterior estimates of the mean and the variance for the QTL effects.

https://doi.org/10.1017/S0016672305007871 Published online by Cambridge University Press

Gaffney, P. (2001). An efficient reversible jump Markov chain Monte Carlo approach to detect multiple loci and their effects in inbred crosses. PhD thesis, University of Wisconsin at Madison, Madison, WI, USA.

Gessler, D. & Xu, S. (1999). Multipoint genetic mapping of quantitative trait loci with dominant markers in outbred populations. *Genetica* **105**, 281–291.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics*, pp. 169–194. New York: Oxford University Press.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Guex, G., Hotz, H. & Semlitsch, R. (2002). Deleterious alleles and differential viability in progeny of natural hemiclonal frogs. *Evolution* **56**, 1036–1044.

Haley, C. & Knott, S. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.

Hawthorne, D. & Via, S. (2001). Genetic linkage and ecological specialization and reproductive isolation in pea aphids. *Nature* **412**, 904–907.

Heath, S. (1997). Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *American Journal of Human Genetics* **61**, 748–760.

Hoeschele, I. (2001). Mapping quantitative trait loci in outbred pedigrees. In *Handbook of Statistical Genetics*, pp. 599–644. New York: Wiley.

Hoeschele, I., Uimari, P., Grignola, F., Zhang, Q. & Gage, K. (1997). Advances in statistical methods to map quantitative trait loci in outbred populations. *Genetics* **147**, 1445–1457.

Hoti, F., Sillanpää, M. & Holmström, L. (2002). A note on estimating the posterior density of a quantitative trait locus from a Markov chain Monte Carlo sample. *Genetic Epidemiology* **22**, 369–376.

Jannink, J. & Fernando, R. (2004). On the Metropolis–Hastings acceptance probability to add or drop a quantitative trait locus in Markov chain Monte Carlo-based Bayesian analysis. *Genetics* **166**, 641–643.

Jansen, R. (1993). Interval mapping of multiple quantitative trait loci. *Genetics* **135**, 205–211.

Knapp, S., Bridges, W. & Birkes, D. (1990). Mapping quantitative trait loci using molecular marker linkage maps. *Theoretical and Applied Genetics* **79**, 583–592.

Lander, E. & Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Luo, L. & Xu, S. (2003). Mapping viability loci using molecular markers. *Heredity* **90**, 459–467.

Maliepaard, C., Sillanpää, M., van Ooijen, J., Jansen, R. & Arjas, E. (2001). Bayesian versus frequentist analysis of multiple quantitative trait loci with an application to an outbred apple cross. *Theoretical and Applied Genetics* **103**, 1243–1253.

Martinez, O. & Curnow, R. (1992). Estimating the locations and the size of the effects of quantitative trait loci using flanking markers. *Theoretical and Applied Genetics* **85**, 480–488.

Martinez, V. (2003). Quantitative genetic analysis using molecular markers with applications to fish populations. PhD thesis, University of Edinburgh, Edinburgh, UK.

Martinez, V., Hill, W. G. & Knott, S. (2002*a*). On the use of double haploids for detecting QTL in outbred populations. *Heredity* **88**, 423–431.

Martinez, V., Sillanpää, M., Thorgaard, G., Robinson, B., Woolliams, J. & Knott, S. (2002*b*). Evidence of a pleiotropic QTL influencing components of early development in double haploid lines of rainbow trout. *7th World Congress on Genetics Applied to Livestock Production*, CD ROM communication 06-08.

McCune, A., Fuller, R., Aquilina, A., Dawley, R., Fadool, J., Houle, D., *et al.* (2002). A low genomic number of recessive lethals in natural populations of bluefin killifish and zebrafish. *Science* **296**, 2398–2401.

Narita, A. & Sasaki, Y. (2004). Detection of multiple QTL with epistatic effects under a mixed inheritance model in an outbred population. *Genetics, Selection and Evolution* **36**, 415–433.

Palti, Y., Shirak, A., Cnaani, G., Hulata, G., Avtalion, R. & Ron, M. (2002). Detection of genes with deleterious alleles in an inbred line of tilapia (*Oreochromis aureus*). *Aquaculture* **206**, 151–164.

Richardson, G. W. & Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. New York: Chapman and Hall.

Ritland, K. (1996). Inferring the genetic basis of inbreeding depression in plants. *Genome* **39**, 1–8.

Robison, B., Wheeler, P. & Thorgaard, G. (1999). Variation in development rate among clonal lines of rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* **173**, 131–141.

Robison, B., Wheeler, P., Sundin, K. & Thorgaard, G. (2001). Composite interval mapping reveals a QTL of major effect on embryonic development rate in rainbow trout (*Onchorhynchus mykiss*). *Journal of Heredity* **92**, 16–22.

Seaman, S., Richardson, S., Stücker, I. & Benhamou, S. (2002). A Bayesian partition model for case-control studies on highly polymorphic candidate genes. *Genetic Epidemiology* **22**, 356–368.

Shoemaker, J., Painter, I. & Weir, B. (1999). Bayesian statistics in Genetics: a guide for the uninitiated. *Trends in Genetics* **15**, 354–358.

Sillanpää, M. & Arjas, E. (1998). Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* **148**, 1373–1388.

Sillanpää, M. & Corander, J. (2002). Model choice in gene mapping: what and why. *Trends in Genetics* **18**, 301–307.

Sillanpää, M., Gasbarra, D. & Arjas, E. (2004). Comment on the 'On the Metropolis–Hastings acceptance probability to add or drop a quantitative trait locus in Markov chain Monte Carlo-based Bayesian analysis'. *Genetics* **167**, 1037–1037.

Smith, B. (2003). *Bayesian Output Analysis* (*BOA*). Manual. Version 1.01. Iowa City: The University of Iowa, College of Public Health.

Sorensen, D. & Gianola, D. (2002). *Likelihood, Bayesian, and McMC Methods in Quantitative Genetics*. Berlin: Springer.

Tanner, M. (1996). *Tools for Statistical Inference. Methods for Exploration of Posterior Distributions and Likelihood Functions*. Berlin: Springer.

Uimari, P. & Sillanpää, M. (2001). Bayesian oligogenic analysis of quantitative and qualitative traits in general pedigrees. *Genetic Epidemiology* **21**, 224–242.

Vogl, C. & Xu, S. (2000). Multipoint mapping of viability and segregation distorting loci using molecular markers. *Genetics* **155**, 1439–1447.

Vos, P., Hogers, M., Bleeker, M., Rijans, M., der Lee, T. V., Hornes, M., *et al.* (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407–4414.

Waagepetersen, R. & Sorensen, D. (2001). A tutorial on reversible jump McMC with a view toward applications

https://doi.org/10.1017/S0016672305007871 Published online by Cambridge University Press

in QTL mapping. *International Statistical Review* **69**, 49–61.

Wang, H., Zhang, Y.-M., Li, X., Masinde, G., Mohan, S., Baylink, D. & Xu, S. (2005). Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics* **170**, 465–480.

Whittaker, J., Thompson, R. & Visscher, P. (1996). On the mapping of QTL by regression of phenotype on marker-type. *Heredity* **77**, 23–32.

Xu, S. (2003). Estimating polygenic effects using markers of the entire genome. *Genetics* **163**, 789–801.

Yi, N. & Xu, S. (1999). Bayesian mapping of quantitative trait loci for complex binary traits. *Genetics* **155**, 1391–1403.

Yi, N. & Xu, S. (2002). Mapping quantitative trait loci with epistatic effects. *Genetical Research* **79**, 185–198.

Yi, N., Xu, S. & Allison, D. (2003). Bayesian model choice and search strategies for mapping interacting quantitative trait loci. *Genetics* **165**, 867–883.

Young, W., Wheeler, P., Coryell, V., Keim, P. & Thorgaard, G. (1998). A detailed linkage map of rainbow trout produced using doubled haploids. *Genetics* **148**, 839–850.

Zeng, Z. (1994). Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468.

Zhang, Y.-M. & Xu, S. (2005). A penalized maximum likelihood method for estimating epistatic effects of QTL. *Heredity* **95**, 96–104.

https://doi.org/10.1017/S0016672305007871 Published online by Cambridge University Press