Heredity (2011) 106, 357–369
© 2011 Macmillan Publishers Limited All rights reserved 0018-067X/11
www.nature.com/hdy

ORIGINAL ARTICLE

Mapping quantitative trait loci using the MCMC procedure in SAS

S Xu and Z Hu
Department of Botany and Plant Sciences, University of California, Riverside, CA, USA

The MCMC procedure in SAS (called PROC MCMC) is particularly designed for Bayesian analysis using the Markov chain Monte Carlo (MCMC) algorithm. The program is sufficiently general to handle very complicated statistical models and arbitrary prior distributions. This study introduces the SAS/MCMC procedure and demonstrates the application of the program to quantitative trait locus (QTL) mapping. A real life QTL mapping experiment in wheat female fertility trait was used as an example for the demonstration. The fertility trait phenotypes were described under three different models: (1) the Poisson model, (2) the Bernoulli model and (3) the zero-truncated Poisson model. One QTL was identified on the second chromosome. This QTL appears to control the switch of seed-producing ability of female plants but does not affect the number of seeds produced once the switch is turned on.

Heredity (2011) **106,** 357–369; doi:10.1038/hdy.2010.77; published online 16 June 2010

Keywords: Bayes; Markov chain Monte Carlo; quantitative trait locus; SAS

Introduction

Most traits of agricultural importance are quantitative in nature (Falconer and Mackay, 1996), for example, grain yield and protein content in corn (Dudley and Johnson, 2009). Many clinical traits in human are also quantitative, for example, obesity and hypertension (Baima *et al*., 1999; Rankinen *et al*., 2006). The current molecular technology allows quick development of linkage map for a species with saturated molecular markers, especially the single-nucleotide polymorphism markers. These markers provide anchors to locate quantitative trait loci (QTL), called QTL mapping. Statistical methods have been well developed for QTL mapping (Lander and Botstein, 1989). However, these methods were mainly for traits with a continuous distribution. When a trait has a discrete distribution, for example, binary disease trait, new methods are required (Xu and Atchley, 1996). Many agriculturally important traits do have discrete distributions. Discrete distribution is also common in human clinical traits, for example, cancer susceptibility (Balmain, 2002). These traits, although simple phenotypically, often have a polygenic background. QTL mapping is important in understanding the genetic architecture for these traits.

Bayesian methods of QTL mapping (Yi and Xu, 2000; Xu *et al*., 2008) are preferable for traits with non-normal distribution, especially when multiple QTL are considered. There are many statistical software packages that can perform Bayesian analysis using the Markov chain Monte Carlo (MCMC) algorithm. Most of the software packages are specialized for particular problems in some special areas. The WinBUGS program, developed by the software development staff of MRC Biostatistics Unit, Cambridge, UK, is perhaps the most popular Bayesian analysis program. The program can be downloaded from the following website, http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml. This program is not problem specific and can handle models with high level of complexity. WinBUGS has been applied to QTL mapping in many studies (Sillanpää and Bhattacharjee, 2005; Yi and Xu, 2008). Another general purpose of Bayesian analysis program is the MCMC procedure in SAS (Chen, 2009; SAS Institute Inc, 2009). The MCMC program is a new procedure in the SAS v9.2 release and is still an experimental version. The procedure has not been as popular as the WinBUGS program because of the short time after the first release of the trial edition. However, because of the wide application of the Bayesian analysis in modern statistics, the MCMC procedure will soon become the most popular Bayesian analysis tool, especially in areas such as statistical genomics and bioinformatics. The MCMC procedure is extremely flexible and general because it can deal with arbitrary prior distributions for the parameters and arbitrary probability density for the data. The only restriction is that the density and the priors must be programmable using the SAS data step functions. This implies that PROC MCMC can handle improper priors. Unlike the WinBUGS program that uses the Gibbs sampler (Geman and Geman, 1984) as default to draw variables, the MCMC procedure in SAS, by default, draws variables using the adaptive block random walk Metropolis algorithm with a normal proposal distribution (Gilks, 2003). The block random walk Metropolis algorithm is an improved version of the general Metropolis-Hastings algorithm (Metropolis *et al*., 1953;

Correspondence: Dr S Xu, Department of Botany and Plant Sciences, University of California, 900 University Avenue, Riverside, CA 92521, USA.
E-mail: shizhong.xu@ucr.edu

Received 23 February 2010; revised 10 May 2010; accepted 13 May 2010; published online 16 June 2010

Hastings, 1970) so that it does not require analytic form of the conditional posterior distribution. As a consequence, the MCMC procedure is the most flexible Bayesian analysis software available so far in the world.

Reports of PROC MCMC application to general data analysis are rare. Our literature search failed to show any publication on the application of this procedure to any fields, although our search may not be sufficiently thorough. We are sure that the MCMC procedure in SAS has never been applied to QTL mapping. As Bayesian QTL mapping has been widely accepted by researchers and will become the main stream method of QTL mapping in the foreseeable future (Satagopan *et al.*, 1996; Sillanpää and Arjas, 1998; Wang *et al.*, 2005; Yi and Xu, 2008), PROC MCMC will become one of the most popular QTL mapping programs. This is particularly evident because many clinic traits have complicated densities beyond the normal distribution (Baima *et al.*, 1999; Balmain, 2002; Rankinen *et al.*, 2006) and Bayesian QTL mapping for these traits is difficult to conduct using other programs.

Four other SAS procedures have the ability to perform Gibbs sampler-implemented Bayesian analysis as an option (SAS Institute Inc, 2009). They are the GENMOD procedure, the PHREG procedure, the LIFEREG procedure and the MIXED procedure. These procedures may be used for Bayesian QTL mapping. The models and priors used in the Bayesian analyses, however, must be found in the list of intrinsic models and priors provided by the SAS procedures. Users do not have the freedom to define their own models and priors. More recently, we published a user defined SAS procedure called the QTL procedure (Hu and Xu, 2009). An option in the PROC QTL statement allows users to choose the Bayesian method for QTL mapping. Again, users do not have the option to choose complicated models and arbitrary priors.

We recently carried out a QTL mapping project for binary traits of line crosses using the MCMC procedure and were very excited about the performance of the procedure. We would like to share our experience with all SAS users who intend to conduct QTL mapping studies in the near future. This report provides an example of QTL mapping using the MCMC procedure. Starting from this example, users can modify and customize the code to analyze their own data using the models and priors of their own choice. Interested users should read the PROC MCMC help document for the syntax and details of the MCMC procedure. Readers of this paper should be regular SAS users and are supposed to be knowledgeable in the MCMC implemented Bayesian method.

Experiment

The experiment was conducted by Dou *et al.* (2009). A female sterile line of wheat XND126 and an elite wheat cultivar Gaocheng8901 with normal fertility were crossed for genetic analysis of female sterility measured as the number of seeded spikelets per plant. The parents, their F_1 and F_2 progeny were planted at the Huaian experimental station in China for the 2006–2007 growing season under the normal autumn sowing condition. The mapping population was an F_2 family consisting of 243 individual plants. About 84% of the F_2 progeny had seeded splikelets and the remaining 16% plants did not have any seeds at all. Among the plants with seeded spikelets, the number of seeded spikelets varied from one to as many as 31. The phenotype is the count data point and can be modeled using the Poisson distribution. A total of 28 SSR markers were used in this experiment. These markers covered five chromosomes of the wheat genome with an average genome marker density of 15.5 cM per marker interval. The five chromosomes are only part of the wheat genome. These chromosomes were scanned for QTL of the Poisson trait using the MCMC implemented Bayesian method. The purpose of the QTL mapping was to identify chromosome regions that are associated with the fertility trait. The dependent variable was the Poisson phenotype, whereas the independent variables were numerically coded genotype indicator variables for the part of genome under investigation. We emphasize the advantage of the Bayesian analysis over the classical maximum likelihood method for detecting multiple QTL simultaneously within a single model. To conduct the multiple locus analysis, we placed one pseudo marker in every 5 cM of the genome. This generated 75 pseudo markers for the five chromosomes (see Supplementary Material for the map of the 75 pseudo markers). Therefore, we had a total of 75 model effects, one for each pseudo marker. For each model effect, the numerically coded coefficient was the difference between the conditional probabilities of the two homozygote genotypes. Let A_1 and A_2 be the alleles carried by Gaocheng8901 and XDN128, respectively. Let A_1A_1, A_1A_2 and A_2A_2 be the three genotypes for the kth pseudo marker of the genome in the F_2 family. The numerically coded value for each locus is

$$\begin{aligned} Z_{jk} =& p(G_{jk} = A_1A_1|\text{marker}) \\ & - p(G_{jk} = A_2A_2|\text{marker}) \end{aligned} \tag{1}$$

for $k = 1,\ldots, 75$, where the conditional probability of QTL genotype given marker information was calculated using the multipoint method of Jiang and Zeng (1997). If a pseudo marker happens to overlap with a fully informative marker, the independent variable Z_{jk} would take one of the three values, 1, 0 and -1, respectively, for the three genotypes, A_1A_1, A_1A_2 and A_2A_2. For the purpose of demonstration, we assumed that there is no dominance effect and thus there is only one Z variable for a locus. The map of the 75 pseudo markers, the phenotypic values (Poisson phenotypes) of the 243 plants and the 75 numerically coded independent variables are also provided in the Supplementary Material of this study.

Model

Poisson data

Let $y_j = \{0, 1, 2,\ldots, \infty\}$ be the observed number of seeded spikelets for the jth plant for $j = 1,\ldots, n$ and $n = 243$. Let μ_j be the expected number of seeded spikelets for the jth plant. The Poisson density for the data point is

$$f(y_j|\mu_j) = \frac{\mu_j^{y_j}}{(y_j)!}\exp(-\mu_j) \tag{2}$$

The expectation is connected to the QTL effects through a log-link function (to be described later).

Binary data

In the F_2 family of 243 plants, 39 of them (16% of 243) had no seeds. It is natural to think that there might be a set of QTL controlling the ability of plant to produce seeds (the seed presence trait) and a set of QTL controlling the number of seeded spikelets once the seeds are produced. The seed presence trait is a binary trait, whereas the number of seeds is a Poisson trait. These two traits may be controlled by different sets of QTL. The binary phenotype is denoted by $y_j = 0$ for no seed and $y_j = 1$ for the presence of seed (regardless of how many seeds). Let μ_j be the expectation of the seed presence for the jth plant for $j = 1, \ldots, n$ and $n = 243$. The binary (also called Bernoulli) density for the data point is

$$f(y_j|\mu_j) = \mu_j^{y_j}(1-\mu_j)^{1-y_j} \tag{3}$$

The expectation is connected to the QTL effects through a probit link function (to be described later).

Truncated Poisson data

For the 204 seeded plants (84% of the 243 plants), the number of seeded spikelets varied. We now examine the genetic basis of the seeded spikelets variation using only the 204 plants. The sample size of this sub-population was $n = 204$ now. Let $y_j = \{1, 2, \ldots, \infty\}$ be the observed number of seeded spikelets for the jth plant for $j = 1, \ldots, n$ and $n = 204$. It is a zero-truncated Poisson distribution. Let μ_j be the expected number of seeded spikelets for the jth plant. The zero-truncated Poisson density for the data point is

$$f_0(y_j|\mu_j) = \frac{f(y_j|\mu_j)}{1-\exp(-\mu_j)} = \frac{\mu_j^{y_j}\exp(-\mu_j)}{(y_j)!\left[1-\exp(-\mu_j)\right]} \tag{4}$$

The expectation is connected to the QTL effects through a log-link function to be described in the following section.

Link function

The log link: For the Poisson and truncated Poisson densities, the log-link function was chosen. Let

$$\eta_j = \beta + \sum_{k=1}^{m} Z_{jk}\gamma_k \tag{5}$$

be the linear model for the QTL effects, where $m = 75$ is the number of pseudo markers, β is the intercept, γ_k is the QTL effect of the kth pseudo marker and Z_{jk} is the conditional expectation of the genotype indicator variable defined earlier. The Z variable defined in such a way so that γ_k is equivalent to the additive effect a defined by Falconer and Mackay (1996). This is the additive model because the dominance effects have been ignored in the model. We are interested in estimating the parameter vector $\theta = \{\beta, \gamma_1, \ldots, \gamma_{75}\}$. The relationship between μ_j and η_j is through the log link,

$$\eta_j = \log(\mu_j) \tag{6}$$

More intuitively, the inverse of the log link is

$$\mu_j = \exp(\eta_j) = \exp\left(\beta + \sum_{k=1}^{m} Z_{jk}\gamma_k\right) \tag{7}$$

The probit link: For the binary density, the probit link function was chosen, although the logit-link function is another option. The probit link is described as

$$\eta_j = \mathrm{probit}(\mu_j) = \Phi^{-1}(\mu_j) \tag{8}$$

More intuitively, the inverse of the probit link is the standardized normal cumulative function,

$$\mu_j = \Phi(\eta_j) = \Phi\left(\beta + \sum_{k=1}^{m} Z_{jk}\gamma_k\right) \tag{9}$$

where $\Phi()$ is the standardized cumulative normal distribution.

Prior distribution

There are many prior distributions from which we can choose. We only chose one type of prior for each parameter as an example. The intercept was assigned a flat normal prior, that is,

$$\pi(\beta) = \mathrm{Normal}(\beta|0,\ 10^{15}) \tag{10}$$

This prior is almost identical to $\pi(\beta) = \mathrm{Normal}\ (\beta\,|\,0, \infty) \propto 1$. Each of the QTL (pseudo marker) effect was assigned
a normal prior,

$$\pi(\gamma_k) = \mathrm{Normal}(\gamma_k|0,\ \sigma_k^2) \tag{11}$$

This prior is QTL specific, that is, each QTL has its own prior variance. The variance in the prior was assigned a higher level prior (hierarchical prior),

$$\pi(\sigma_k^2) = \text{Inv-}\chi^2(\gamma_k|\tau,\ \omega) = \text{Inv-}\chi^2(\gamma_k|10^{-10},\ 10^{-10}) \tag{12}$$

This prior is not much different from the Jeffreys' prior (Berger, 1985),

$$\pi(\sigma_k^2) = \text{Inv-}\chi^2(\gamma_k|0,\ 0) = 1/\sigma_k^2 \tag{13}$$

This hierarchical model is also called the Bayesian shrinkage analysis (Wang *et al*., 2005).

SAS code

The SAS codes to read the data, to analyze the data and to report the result are provided in this section. We assumed that the original data are stored in a folder named 'c:\mcmc\fertility' with a file name 'fertility.csv'. The posterior sample is written to the same folder with a file name 'post-sample.csv'. The file locations and the names of the input and output files should be customized by the users.

Poisson data analysis

The SAS code for the Poisson data analysis is given in Table 1. Here are explanations of the SAS code. The statements before 'proc mcmc' are typical SAS statements for data input (creating a SAS data set). Readers are supposed to be familiar with the SAS language, and thus no explanation was given. The MCMC procedure starts with the statement 'proc mcmc'. The SAS code in this example is defined in a macro named 'fertility'.

Table 1 SAS code for the Poisson data analysis

```
%let dir=c:\mcmc\fertility;

libname xx "&dir";
filename aa "&dir\fertility.csv" lrecl=200000;
filename bb "&dir\post-sample.csv";

data fertility;
     infile aa dlm=',' firstobs=2;
     input plant y z1-z75;
run;

%macro fertility;
ods graphics on;
proc mcmc data=fertility outpost=xx.postsample seed=12345
          nmc=50000 thin=50 nbi=5000 simreport=5
          monitor=(beta gamma1-gamma3 sigmasqr1-sigmasqr3)
          stats(percent=(2.5 5 50 95 97.5))=all
          diagnostics=(all geweke(f1=0.3 f2=0.3));
     ods select PostSummaries ESS Geweke PostIntervals TADpanel;
     array z[75];
     array gamma[75];
     parms beta 0;
     prior beta ~ normal(mean=0,var=1e15);
     %do k=1 %to 75;
         parms gamma&k 0;
         parms sigmasqr&k 1;
         prior gamma&k ~ normal(mean=0, var=sigmasqr&k);
         prior sigmasqr&k ~ sichisq(1e-10,1e-10);
     %end;
     eta=beta;
     do k=1 to 75;
        eta=eta+z[k]*gamma[k];
     end;
     mu = exp(eta);
     model y ~ poisson(mu);
run;
ods graphics off;
%mend;

%fertility

proc export data=xx.postsample outfile=bb dbms=csv replace;

run;
```

The reason for using the SAS macro will be given in Appendix A after all the statements of the MCMC procedure are explained. More explanations are also given in Appendix A.

Binary data analysis

The SAS code for the binary data analysis is given in Table 2. The code differs from that of the Poisson data analysis only by a few lines. Explanations for the few extra lines are given in Appendix B.

Truncated Poisson data analysis

The SAS code of the truncated Poisson data analysis is given in Table 3. The explanations for the few statements that differ from the previous SAS codes are given in Appendix C.

Result

MCMC procedure

Each of the three data analyses took about 9 h of central processing unit time (2.5 GHz and 3.25 GB of RAM) to complete the MCMC sampling. The most important output of each analysis was the posterior sample saved in the outpost = data set. From this posterior sample, users can obtain summary statistics about the parameters of interest. In addition to the posterior sample, users can choose to report the summary and diagnostic statistics for the parameters specified in the monitor = () option in

Table 2 The SAS code for the binary data analysis

```
%let dir=c:\mcmc\fertility;

libname xx "&dir";
filename aa "&dir\fertility.csv" lrecl=200000;
filename bb "&dir\post-sample.csv";

data fertility;
     infile aa dlm=',' firstobs=2;
     input plant y z1-z75;
     y = (y>0);
run;

%macro fertility;
ods graphics on;
proc mcmc data=fertility outpost=xx.postsample seed=12345
          nmc=50000 thin=50 nbi=5000 simreport=5
          monitor=(beta gamma1-gamma3 sigmasqr1-sigmasqr3)
          stats(percent=(2.5 5 50 95 97.5))=all
          diagnostics=(all geweke(f1=0.3 f2=0.3));
     ods select PostSummaries ESS Geweke PostIntervals TADpanel;
     array z[75];
     array gamma[75];
     parms beta 0;
     prior beta ~ normal(mean=0,var=1e15);
     %do k=1 %to 75;
         parms gamma&k 0;
         parms sigmasqr&k 1;
         prior gamma&k ~ normal(mean=0, var=sigmasqr&k);
         prior sigmasqr&k ~ sichisq(1e-10,1e-10);
     %end;
     eta=beta;
     do k=1 to 75;
        eta=eta+z[k]*gamma[k];
     end;
     mu = probnorm(eta);
     model y ~ binary(mu);
run;
ods graphics off;
%mend;

%fertility

proc export data=xx.postsample outfile=bb dbms=csv replace;
run;
```

the output window. Table 4 demonstrates the summary statistics table for the variables monitored in the MCMC procedure for the binary data analysis. Figure 1 shows the trace-autocorrelation-density (TAD) panel produced by the MCMC procedure for this binary data analysis.

The 75 pseudo markers were distributed along five chromosomes of the wheat genome. The posterior mean (Bayesian estimate) and the $\alpha = 0.10$ equal-tail credible interval (bracketed by the 5 percentile and 95 percentile) are plotted against the marker location, forming an estimated QTL effect profile and two credible interval profiles (see Figure 2). The top panel of Figure 2a shows the result of the Poisson data analysis. A marker in the second chromosome (at about 100 cM of the genome) shows an association with the Poisson trait. The estimated effect was different from zero with high credibility (the credible interval excluded zero). There appears to be some activity towards the end of chromosome 5, but with very low credibility (the credible interval included zero). Therefore, we are confident that one QTL has been detected for the Poisson trait in chromosome 2.

The QTL effect profile and the credible interval profiles for the binary data analysis are shown in Figure 2b, the panel in the middle. The pseudo marker identified in the Poisson data analysis was also identified for the binary data analysis (the pseudo marker at position 100 cM of the genome). The equal-tail credible interval excluded zero. Although more pseudo markers showed some activities (estimated QTL effects deviating from zero), none of them had high credibility.

Table 3 The SAS code for the zero-truncated Poisson data analysis

```
%let dir=c:\mcmc\fertility;

libname xx "&dir";
filename aa "&dir\fertility.csv" lrecl=200000;
filename bb "&dir\post-sample.csv";

data fertility;
     infile aa dlm=',' firstobs=2;
     input plant y z1-z75;
     if y > 0;
     fy=log(fact(y));
run;

%macro fertility;
ods graphics on;
proc mcmc data=fertility outpost=xx.postsample seed=12345
          nmc=50000 thin=50 nbi=5000 simreport=5
          monitor=(beta gamma1-gamma3 sigmasqr1-sigmasqr3)
          stats(percent=(2.5 5 50 95 97.5))=all
          diagnostics=(all geweke(f1=0.3 f2=0.3));
     ods select PostSummaries ESS Geweke PostIntervals TADpanel;
     array z[75];
     array gamma[75];
     parms beta 0;
     prior beta ~ general(log(1));
     %do k=1 %to 75;
         parms gamma&k 0;
         parms sigmasqr&k 1;
         prior gamma&k ~ normal(mean=0, var=sigmasqr&k);
         prior sigmasqr&k ~ general(-log(sigmasqr&k));
     %end;
     eta=beta;
     do k=1 to 75;
        eta=eta+z[k]*gamma[k];
     end;
     mu = exp(eta);
     f=y*log(mu)-mu-fy;
     g=log(1-exp(-mu));
     f0=f-g;
     model y ~ general(f0);
run;
ods graphics off;
%mend;

%fertility

proc export data=xx.postsample outfile=bb dbms=csv replace;

run;
```

The panel at the bottom of Figure 2c shows the result for the zero-truncated Poisson data analysis. The QTL controlling the Poisson and the binary traits in chromosome 2 was not detectable for the truncated Poisson trait. This means that the QTL in chromosome 2 only controls the switch from seed absence to seed presence. It does not control the numbers of seeds. We can see some activities towards the end of chromosome 5. This time, the credible interval almost excluded zero. Although the credibility was not high, it was stronger than what we saw in the Poisson data analysis. We can claim an association between a pseudo marker in the end of chromosome 5 and the number of seeds, but only with modest credibility. The overall conclusion was that the seed presence trait and the number of seeds are controlled by different sets of QTL.

Table 5 gives the detailed information about the large QTL identified for the Poisson and binary traits and the suggested QTL identified for the truncated Poisson trait. The QTL on chromosome 2 is close to the right hand side marker (0.66 cM away). This explains the high credibility of the QTL. The suggested QTL on chromosome 5

Table 4 Posterior sample summary statistics for the variables monitored in PROC MCMC for the binary data analysis

Parameter	N	*Mean*	*s.d.*	*Percentile*				
				2.50%	*5%*	*50%*	*95%*	*97.50%*
Beta	1000	1.5679	0.1935	1.1965	1.2422	1.5580	1.8951	1.9614
Gamma1	1000	0.1814	0.3704	−0.2814	−0.1665	0.0696	0.8969	1.1676
Gamma2	1000	0.1351	0.4568	−0.9653	−0.5517	0.0670	0.8849	1.0845
Gamma3	1000	0.0139	0.2280	−0.2247	−0.0564	0.000193	0.1876	0.4220
Sigmasqr1	1000	3.4429	12.0789	0.00089	0.0012	0.0438	23.4292	50.0437
Sigmasqr2	1000	1.4469	4.9278	0.00133	0.00133	0.0926	7.225	12.7707
Sigmasqr3	1000	0.4501	3.3120	0.000023	0.000023	0.000045	0.3378	2.5623

is 52 cM away from the left marker and 17 cM away from the right marker. Therefore, there is virtually no information for the large interval, explaining the low credibility (wide credibility interval) for this QTL. The estimated QTL effects were used to calculate the expected values of the traits for different genotypes. Let A_1 be the allele of the fertile parent Gaocheng8901 and A_2 be the allele of the sterile parent XND126. For the Poisson trait, the expected numbers of seeded spikelets for the three genotypes of QTL 1 are

$$\begin{bmatrix}\mu(A_1A_1)\\ \mu(A_1A_2)\\ \mu(A_2A_2)\end{bmatrix}=\begin{bmatrix}\exp(\beta+\gamma)\\ \exp(\beta)\\ \exp(\beta-\gamma)\end{bmatrix}=\begin{bmatrix}\exp(2.9165+0.4089)\\ \exp(2.9165)\\ \exp(2.9165-0.4089)\end{bmatrix}=\begin{bmatrix}27.8101\\ 18.4765\\ 12.2754\end{bmatrix} \tag{14}$$

For the binary trait, the expected probabilities of seed producing for the three genotypes of QTL 1 are

$$\begin{bmatrix}\mu(A_1A_1)\\ \mu(A_1A_2)\\ \mu(A_2A_2)\end{bmatrix}=\begin{bmatrix}\Phi(\beta+\gamma)\\ \Phi(\beta)\\ \Phi(\beta-\gamma)\end{bmatrix}=\begin{bmatrix}\Phi(1.5679+1.5867)\\ \Phi(1.5679)\\ \Phi(1.5679-1.5867)\end{bmatrix}=\begin{bmatrix}0.9992\\ 0.9415\\ 0.4925\end{bmatrix} \tag{15}$$

For the truncated Poisson trait, the expected numbers of seeded spikelets for the three genotypes of QTL 2 are

$$\begin{bmatrix}\mu(A_1A_1)\\ \mu(A_1A_2)\\ \mu(A_2A_2)\end{bmatrix}=\begin{bmatrix}\exp(\beta+\gamma)\\ \exp(\beta)\\ \exp(\beta-\gamma)\end{bmatrix}=\begin{bmatrix}\exp(3.1144+0.6715)\\ \exp(3.1144)\\ \exp(3.1144-0.6715)\end{bmatrix}=\begin{bmatrix}44.0753\\ 22.5199\\ 11.5064\end{bmatrix} \tag{16}$$

Interval mapping

To compare the MCMC analysis with existing methods, we also performed interval mapping for the female fertility trait under the maximum likelihood framework. In the interval mapping, we scanned the whole genome with a 5 cM increment for a total of 75 putative positions (pseudo markers). The model contained one putative QTL at a time. The entire genome scanning required 75 separate analyses, one for each putative position. We used the GENMOD procedure in SAS to perform the interval mapping. PROC GENMOD can handle Poisson data with the log-link function and binary data with the probit (or logit)-link function. The truncated Poisson data analysis with the GENMOD procedure requires a user-defined density and link function, which we have not figured out yet, and thus we simply deleted the observations with zero seeds from the data and analyzed the remaining 204 plants using the Poisson model.

Figure 3 shows the estimated QTL effects plotted against the genome location for the Poisson data (a), the binary data (b) and the zero-truncated Poisson data (c). For the Poisson data analysis, the QTL effect was quite large across the entire chromosome 2. Chromosomes 1 and 3 also showed some effects, although not as large as chromosome 2. The binary data analysis generated QTL effect profiles with almost the same pattern as the Poisson data analysis, except that the confidence intervals are wider. The zero-truncated Poisson data analysis showed no evidence of QTL effects across the entire genome.

Figure 4 shows the LOD score profiles of the three data analyses. Using the permutation analysis (Churchill and Doerge, 1994), we generated genome-wide critical values for the LOD scores. The threshold values were 11.96, 2.75 and 2.44, respectively, for the Poisson data, the binary data and the truncated Poisson data. Interval mapping detected one QTL at approximately the same position as the one detected in the MCMC analysis. However, the signals of the LOD test statistic profiles for the interval mapping are not as sharp as those of the MCMC procedure. The entire chromosome 2 has LOD scores greater than the critical values for the Poisson and binary data analysis. The truncated Poisson data analysis showed no QTL across the whole genome.

In summary, the interval mapping detected the entire chromosome 2 as significant but the MCMC analysis narrowed down a QTL to a single pseudo marker near the end of chromosome 2. Therefore, the MCMC analysis outperformed the maximum likelihood interval mapping.

Discussion

We used the wheat female fertility trait as an example to demonstrate the MCMC procedure for QTL mapping. We chose the simplest additive model for the presenta-

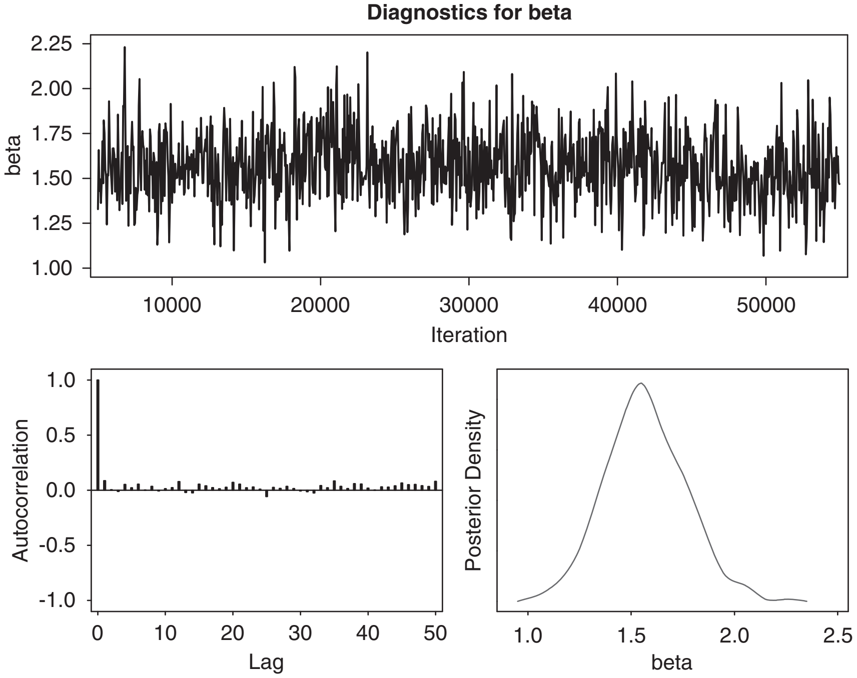

Figure 1 Convergence diagnostics for the population mean (intercept) of the binary data analysis. This diagnostic analysis is represented by the trace-autocorrelation-density (TAD) panel in the MCMC procedure of SAS.

tion. Dominance effects have been ignored for simplicity. The mapping population was small ($n = 243$) and the marker density was low (15.5 cM per interval). A more complete analysis should be conducted with denser marker map in large population. In addition, dominance and epistatic effects should also be considered in the complete analysis. The MCMC procedure can handle multiple QTL with dominance and epistatic effects. For the complete analysis, each locus should have two genotype indicator variables, one for the additive effect and one for the dominance effect. Each pair of loci should have four epistatic effects, additive by additive, additive by dominance, dominance by additive and dominance by dominance. This study emphasizes the MCMC procedure, not the biology, and thus a complete analysis was not conducted.

Another simplification we made here was the definition of the Z variables. We took the Haley and Knott (1992) approach by substituting the missing QTL genotypes by the conditional expectations given marker information. In a fully Bayesian analysis, we should sample the Z_{jk} variables from the conditional posterior distributions. Let

$$\delta_j = [\delta_{j1} \quad \delta_{j2} \quad \delta_{j3}] \tag{17}$$

be a multinomial variable with sample size one (multivariable Bernoulli variable) taking value [1 0 0], [0 1 0] or [0 0 1], respectively, for A_1A_1, A_1A_2 or A_2A_2. One can calculate the conditional posterior distribution of δ_j (see Wang *et al.*, 2005), denoted by

$$p_j^* = [\Pr(\delta_{j1} = 1|\cdots) \quad \Pr(\delta_{j2} = 1|\cdots) \quad \Pr(\delta_{j3} = 1|\cdots)] \tag{18}$$

The multinomial variable δ_j can be sampled from

$$p(\delta_j|\cdots) = \text{Multinomial}(\delta_j|1, p_j^*) \tag{19}$$

Once δ_j is sampled, the Z_{jk} variable simply takes

$$Z_{jk} = \delta_{j1} - \delta_{j3} \tag{20}$$

If dominance effects are considered in the model, a W variable is needed to capture the dominance effect for each locus. The W variable is defined as

$$W_{jk} = \delta_{j2} - (\delta_{j1} + \delta_{j3}) \tag{21}$$

The Haley and Knott's (1992) definitions of the Z and W variables simply take the expectations

$$Z_{jk} = E(\delta_{j1} - \delta_{j3}) = E(\delta_{j1}) - E(\delta_{j3}) \tag{22}$$

and

$$W_{jk} = E(\delta_{j2}) - [E(\delta_{j1}) + E(\delta_{j3})] \tag{23}$$

These expectations are then treated as known values throughout the analysis.

The MCMC procedure in SAS can handle very complicated models. The Poisson and binary data analyses were already more complicated than the continuously distributed normal traits. The truncated Poisson trait was even more complicated. Although PROC MCMC can handle truncated Poisson density, we chose to define the density in the programming statements to show the flexibility of the MCMC procedure. We can see that the MCMC procedure only requires very limited recoding to accomplish the complicated models. This property of the MCMC procedure is unique and no other Bayesian programs can compete with it.

An alternative way to handle this type of data is to use the zero-inflated Poisson model (Cui and Yang, 2009). We can fit two models to the same data simultaneously.

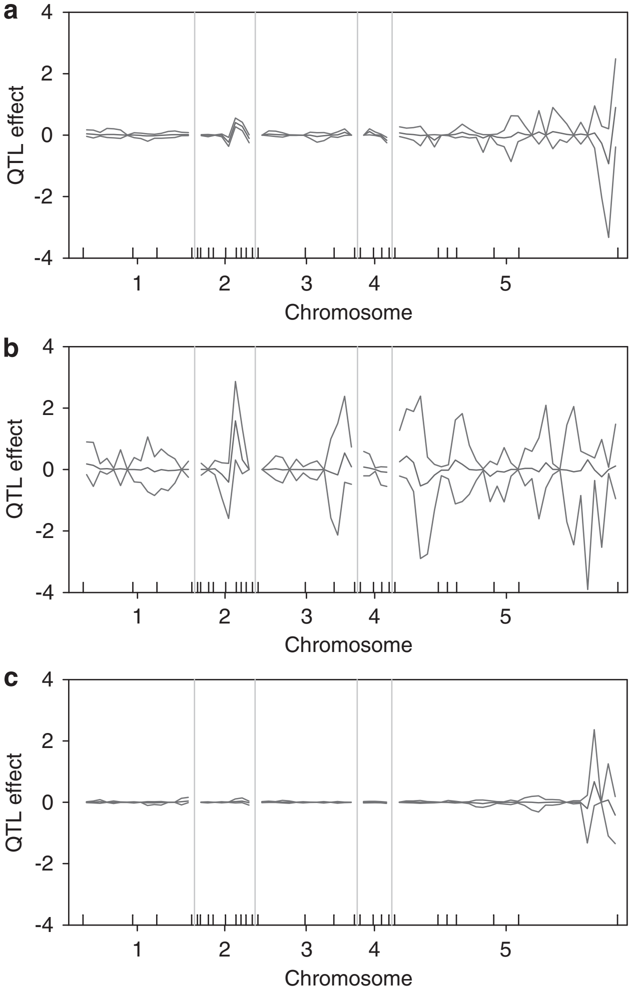

Figure 2 Bayesian estimate of QTL effect for the female fertility trait of wheat. The blue curve is the profile of the posterior mean of the QTL effect. The red lines define $\alpha=0.10$ equal-tail credible interval (5 and 95 percentiles) of the posterior distribution of the QTL effect. The five chromosomes are separated by four vertical references lines. The marker positions are indicated by the ticks on the horizontal axis. (**a**) The top panel gives the result for the Poisson data analysis with sample size $n=243$. (**b**) The panel in the middle shows the result for the binary data analysis with sample size $n=243$. (**c**) The panel at the bottom is the result of the zero-truncated Poisson data analysis with sample size $n=204$. A full color version of this figure is available at the *Heredity* Journal online.

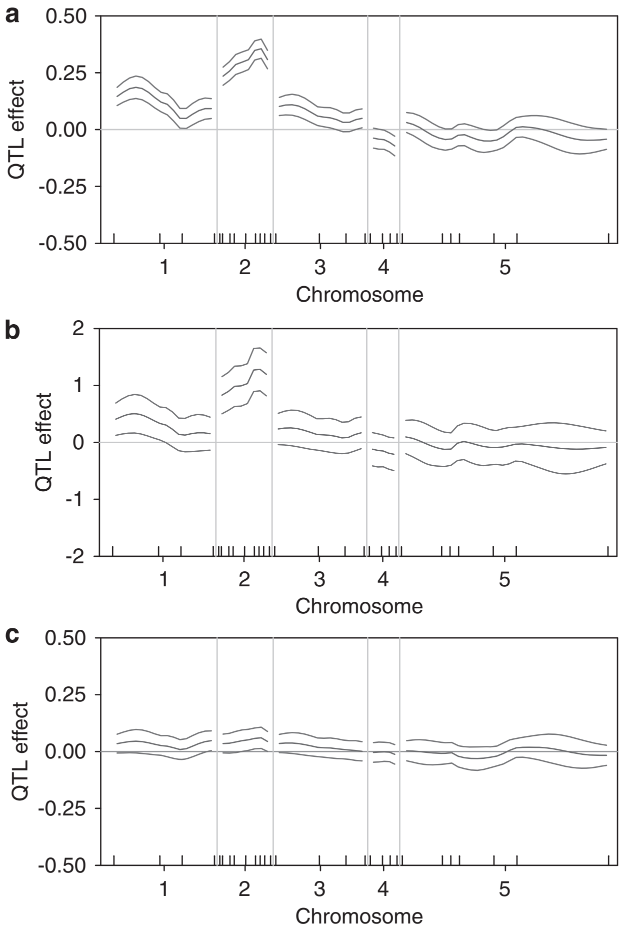

Figure 3 Maximum likelihood estimate of QTL effect for the female fertility trait of wheat. The blue curve is the profile of the estimated QTL effect. The red lines define the $\alpha=0.10$ confidence interval (5 and 95 percentiles) of the QTL effect. The five chromosomes are separated by four vertical references lines. The marker positions are indicated by the ticks on the horizontal axis. (**a**) The top panel gives the result for the Poisson data analysis with sample size $n=243$. (**b**) The panel in the middle shows the result for the binary data analysis with sample size $n=243$. (**c**) The panel at the bottom is the result of the zero-truncated Poisson data analysis with sample size $n=204$. A full color version of this figure is available at the *Heredity* Journal online.

Table 5 Information about the identified QTL by the MCMC procedure for the three data analyses (Poisson, binary and truncated Poisson)

	QTL 1	*QTL 2*	*Intercept*
Chromosome	2	5	
Left marker	Xgwm296 (19.63 cM)	Xwmc291 (86.54 cM)	
QTL location	26.33 cM	138.63 cM	
Right marker	Xbarc95 (26.97 cM)	cft21 (155.66 cM)	
Credibility	High	Low	
Effect (Poisson)	0.4089 (0.0859)	—	2.9165 (0.0210)
Effect (binary)	1.5867 (0.7337)	—	1.5679 (0.1935)
Effect (trunc-Poi)	—	0.6715 (0.9169)	3.1144 (0.0182)

Abbreviations: MCMC, Markov chain Monte Carlo; QTL, quantitative trait locus.
The posterior means and posterior s.d.s (in parentheses) for the QTL effects and the intercepts are presented in the last three rows of the table.

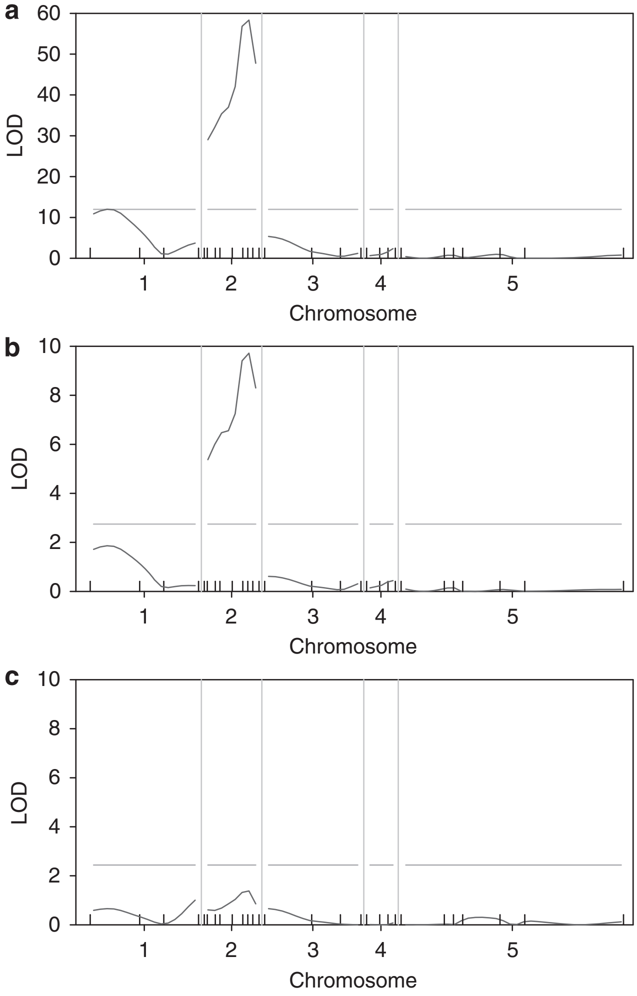

Figure 4 LOD scores for the maximum likelihood method of QTL mapping for the female fertility trait of wheat. The horizontal straight line represents permutation-generated threshold value for the LOD score test statistic. The five chromosomes are separated by four vertical references lines. The marker positions are indicated by the ticks on the horizontal axis. (**a**) The top panel gives the result for the Poisson data analysis with sample size $n = 243$. (**b**) The panel in the middle shows the result for the binary data analysis with sample size $n = 243$. (**c**) The panel at the bottom is the result of the zero-truncated Poisson data analysis with sample size $n = 204$.

One is the Poisson model and the other is the 'zero model'. The zero model captures QTL for the binary trait and the Poisson model captures QTL responsible for the variation of the number of seeded spikelets. The MCMC procedure can be used to handle the zero-inflated Poisson model by adding a user defined log-likelihood function. We should also describe the regression coefficients using a multivariate normal prior for both the model effect and the zero model effect jointly. The additional coding requires handling matrix algebra, which can be done but a little tedious, and we have not figured it out at this moment. We will prioritize the zero-inflated Poisson model as our next project. At this moment, the three different models are sufficient to demonstrate the usefulness of the procedure.

Unfortunately, the high flexibility of PROC MCMC is traded off by the low computational efficiency in terms of long time taken for completing the MCMC sampling process. The MCMC procedure is extremely time consuming for large models (models with a large number of parameters). It is highly efficient for small but complicated models. By default, the MCMC procedure samples all variables using the block random walk Metropolis algorithm. This algorithm requires tuning the parameters of the proposal distribution. Most of the central processing unit time is actually taken for tuning the proposal distribution. This explains the low computational efficiency. There is an option in the 'proc mcmc' statement that allows users to skip the tuning step. This option is to set the number of iterations of each tuning loop to zero, that is, 'ntu = 0'. However, without the tuning process, the MCMC sampler may requires an even longer chain to reach the stationary distribution. Therefore, it is not advised to skip the tuning step. The MCMC procedure does provide an option for skilled SAS users to write their own samplers for the parameters of interest. This option is called user-defined sampler (UDS). If the sampler for a parameter is UDS, the tuning process is not needed. If the UDS is a Gibbs sampler, the acceptance rate is 100% (most efficient). However, writing UDSs can be very cumbersome and, by doing that, we are not taking advantage of the MCMC procedure. In addition, the overhead cost of calling other procedures when UDS is used may further slow down the speed. As the MCMC procedure is an experimental procedure, much improvement in terms of the computing speed is expected in future releases.

The SAS/MCMC procedure is an excellent tool for teaching the Bayesian method of QTL mapping. As more researchers and students are interested in the Bayesian method, a user friendly software package is necessary and the SAS/MCMC procedure can fulfill that need. The data input and output are all handled by SAS within the same environment. Students only need to learn the MCMC statements, which are simple and easy to code.

In terms of Bayesian QTL mapping, existing programs are available, for example, Bqtl (Bayesian QTL mapping, The R Development Core Team, 2001), MultiMapper (Sillanpää and Arjas, 1998), R/qtlbim (Bayesian interval mapping, Yandell *et al*., 2007). They are more efficient than the MCMC procedure. We recently released a user-defined SAS procedure, the QTL procedure (Hu and Xu, 2009). With the QTL procedure in SAS, users can choose the method = 'bayes' option in the proc QTL statement. This option will turn on the MCMC implemented Bayesian shrinkage algorithm for QTL mapping (Xu, 2003; Wang *et al*., 2005). Unfortunately, users have no freedom to choose their own priors and likelihood. The high computational efficiency for the specialized programs is compromised by the low flexibility. The MCMC procedure for QTL mapping can be very efficient if users only want to investigate a target region of the genome for an extremely complicated likelihood for the trait and complicated priors for parameters. The truncated Poisson trait is an example of complicated likelihood. The scaled inverse χ^2 distribution for the prior of a variance component is not very complicated, but one can easily handle hierarchical models by assigning prior distributions to the degree of belief and scale parameter in the scaled inverse χ^2 distribution. Details of the hierarchical Bayesian model for QTL mapping can be found in the study by Yi and Xu (2008).

Finally, the map of the 75 pseudo markers, the original data (plant id, phenotype and numerically coded genotypes for the 75 markers) and the SAS codes presented in the main text of this article can be downloaded from our personal website (http://www.statgen.ucr.edu) under the PROC MCMC software section. Readers can use the data and the codes for practicing and testing the MCMC procedure.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This project was supported by the Agriculture and Food Research Initiative (AFRI) of the USDA National Institute of Food and Agriculture under the Plant Genome, Genetics and Breeding Program 2007-35300-18285.

References

Baima J, Nicolaou M, Schwartz F, DeStefano AL, Manolis A, Gavras I *et al.* (1999). Evidence for linkage between essential hypertension and a putative locus on human chromosome 17. *Hypertension* **34**: 4–7.

Balmain A (2002). Cancer as a complex genetic trait: tumor susceptibility in humans and mouse models. *Cell* **108**: 145–152.

Berger J (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. Springer Verlag: New York.

Chen F (2009). *SAS Global Forum 2009*. Inc SI (ed.). SAS Institute Inc.: Cary, NC.

Churchill GA, Doerge RW (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–971.

Cui Y, Yang W (2009). Zero-inflated generalized Poisson regression mixture model for mapping quantitative trait loci underlying count trait with many zeros. *J Theor Biol* **256**: 276–285.

Dou B, Hou B, Xu H, Lou X, Chi X, Yang J *et al.* (2009). Efficient mapping of a female sterile gene in wheat (*Triticum aestivum L.*). *Genet Res* **91**: 337–343.

Dudley JW, Johnson GR (2009). Epistatic models improve prediction of performance in corn. *Crop Sci* **49**: 763–770.

Falconer DS, Mackay TFC (1996). *Introduction to Quantitative Genetics*, 4th edn. Addison Wesley Longman: Harlow, Essex, UK.

Geman S, Geman D (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* **6**: 721–741.

Gilks W (2003). *Software from MRC Biostatistics Unit*. MRC Biostatistics Unit: Cambridge, UK.

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

Hastings WK (1970). Monte Carlo sampling method using Markov chains and their applications. *Biometrika* **57**: 97–109.

Hu Z, Xu S (2009). PROC QTL—A SAS procedure for mapping quantitative trait loci. *Int J Plant Genomics* **2009**: 141234.

Jiang C, Zeng ZB (1997). Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica* **101**: 47–58.

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

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953). Equation of state calculation by fast computing machines. *J Chem Phys* **21**: 1087–1092.

Rankinen T, Zuberi A, Chagnon YC, Weisnagel SJ, Argyropoulos G, Walts B *et al.* (2006). The human obesity gene map: the 2005 update. *Obesity* **14**: 529–644.

SAS Institute Inc (2009). *The MCMC Procedure, SAS/STAT Help Documentation*. SAS Institute Inc.: Cary, NC.

Satagopan JM, Yandell BS, Newton MA, Osborn TC (1996). A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* **144**: 805–816.

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

Sillanpää MJ, Bhattacharjee M (2005). Bayesian association-based fine mapping in small chromosomal segments. *Genetics* **169**: 427–439.

The R Development Core Team (2001). *BQTL (Bayesian Quantitative Trait Locus Mapping)*. MRC Biostatistics Unit: Cambridge, UK.

Wang H, Zhang Y, Li X, Masinde GL, Mohan S, Baylink DJ *et al.* (2005). Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics* **170**: 465–480.

Xu C, Wang X, Li Z, Xu S (2008). Mapping QTL for multiple traits using Bayesian statistics. *Genet Res* **90**: 1–15.

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

Xu S, Atchley WR (1996). Mapping quantitative trait loci for complex binary diseases using line crosses. *Genetics* **143**: 1417–1424.

Yandell BS, Mehta T, Banerjee S, Shriner D, Venkataraman R, Moon JY *et al.* (2007). R/qtlbim: QTL with Bayesian interval mapping in experimental crosses. *Bioinformatics* **23**: 641–643.

Yi N, Xu S (2000). Bayesian mapping of quantitative trait loci for complex binary traits. *Genetics* **155**: 1391–1403.

Yi N, Xu S (2008). Bayesian LASSO for quantitative trait loci mapping. *Genetics* **179**: 1045–1055.

Appendix A

Detailed explanations of the SAS code for the Poisson data analysis

```
%macro fertility; - Create a SAS macro named 'fertility'.
ods graphics on; - Turn the graphics Output Delivery System (ODS) on.
proc mcmc - Call the MCMC procedure
data=fertility - This is an option of the proc mcmc statement. It tells proc mcmc to use data with a name
        fertility.
outpost=xx.postsample - Tells proc mcmc to write the posterior sample to a SAS dataset named
        xx.postsample. The two level SAS dataset name means that the posterior sample will be stored in
        the folder with libname xx as a permanent SAS dataset. The dataset contains all variables defined
        in the parms and prior statements plus the log likelihood, the log prior density and the log-
        posterior density. If users define some functions of the variables, these functions will also be
        stored in the posterior sample dataset.
```

seed=12345 - This option allows users to set the seed for random number generators. Choosing the same seed will allow the users to duplicate the results. If no seed is given, proc mcmc will assumes a default seed of zero, which will generate a different sequence of random numbers every time the program is executed.

nmc=50 000 - This option defines the total length of the Markov chain excluding the burn-in deletion.

thin=50 - This option defines the thinning rate. In this case, the posterior sample will keep one draw in every 50 iterations. In this example, the posterior sample size (named xx.postsample) will contain 50 000/50=1000 observations.

nbi=5000 - Defines the number of iterations in the burn-in period. In this case, proc mcmc starts to collect posterior sample after 5000 iterations. The burn-in period does not affect the posterior sample size stored in the outpost dataset. For example, the current setting requires proc mcmc to run a total of 50 000+5000=55 000 iterations, although the posterior sample only contains 1000 observations.

simreport=5 - This option tells proc mcmc to report the progress of the MCMC sampling. It is useful for running a large model that takes a very long computing time. The procedure will write a message on the SAS log window 5 times during the MCMC sampling process to tell the user how much time left for the program to finish. Note that proc mcmc only starts to report the progress after the tuning period ends. The tuning time can vary from data to data. For a large model, the tuning time may be longer than the sampling time. For example, for some data, proc mcmc may take 20 h for tuning and 10 h for sampling. If you set simreport=5, the program starts to report the progress when the sampling process starts (after 20 h) and report the progress in every 2 h (2 × 5=10 h) until the sampling progress finishes.

monitor=(beta gamma1-gamma3 sigmasqr1-sigmasqr3) - Variables included in the braces will be subject to post MCMC analysis. Note that there are 75 gamma's and 75 sigmasqr's. We only monitor the first three gamma's and the first three sigmasqr's.

stats(percent=(2.5 5 50 95 97.5))=all - Tell the program to report the percentile values defined in the percent=() option for all the variables included in the monitor=() option.

diagnostics=(all geweke(f1 =0.3 f2 =0.3)) - Tells the program to report the Geweve *z*-test convergence diagnose statistics using the first 30% of the posterior sample and the last 30% of the posterior sample for all the variables included in the monitor=() option.

ods select PostSummaries ESS Geweke PostIntervals TADpanel; - This statement tells proc mcmc to select the following items to be handled by the SAS output delivery system (ODS) for output: (1) The post MCMC summaries for the variables contained in the monitor=() option, (2) the effective sample sizes, (3) the Geweke z-test diagnostic statistics for convergence, (4) the credibility intervals and (5) the trace-autocorrelation-density (TAD) panels. Each monitored variable has a TAD panel that contains three figures drawn in the same page (the trace plot, the autocorrelation plot and the marginal posterior density).

array z[75]; - Define an array named z which refers to z1—z75.

array gamma[75]; - Define an array named gamma. Later on, you can refer gamma1-gamma75 for the 75 variables defined by this array statement. Note the difference between array gamma {75} usually defined in the data step and array gamma[75] defined here.

parms beta 0; - Define a parameter named beta and assign a value 0 as the initial value.

%do k =1 %to 75; - Starts a do-loop 75 times.

parms gamma&k 0; - Define parameter gamma[k] and initialize with 0.

parms sigmasqr&k 1; - Define parameter sigmasqr[k] and initialize with 1.

prior gamma&k ~ normal(mean=0, var=sigmasqr&k);

prior sigmasqr&k ~ sichisq(1e-10,1e-10);

%end; - Ends the *do-loop*.

The MCMC procedure does not allow users to define the parms variables and their priors using the notation gamma[k]. This explains the use of the SAS macro.

eta=beta; - Assign eta the value of beta.

do k =1 to 75; - Define a do-loop.

eta=eta+*z*[k]*gamma[k]; - The use of gamma[k] is legal here in the assign statement, although it is not legal in the parms and prior statements.

end;

mu=exp(eta); - Define the log link (inverse is exponential).

model *y* ~ poisson(mu); - Define the Poisson density.

ods graphics off; - Turn off the ODS graphics.

%mend; - Ends the macro.

%fertility - Execute the macro.

proc export data=xx.postsample outfile=bb dbms=csv replace; - Writes the posterior sample stored in the SAS dataset xx.postsample into a physical excel file with a name defined in the filename bb statement. The filename bb refers to a physical file 'post-sample.csv' in the 'c:\mcmc\fertility' folder.

If we want to sample all the 75 gamma variables together as a block, you can use the following statement,

parms gamma: 0;

The notation ‘gamma:’ is a short expression of gamma1-gamma75. It will be a nightmare for the MCMC procedure to sample that many parameters in a block. It will take forever for the program to tune the parameters of the proposal distribution. One can skip the tuning step, but takes a risk of not converging to the stationary distribution for the Markov chain within a reasonable time frame.

Appendix B

Detailed explanations of the SAS code for the binary data analysis

Majority of the statements are the same as the SAS code given in the Poisson data analysis. The binary analysis requires a few different statements, which are explained below.

$y = (y > 0)$; - This statement in the data step redefines variable y as a binary variable.
mu=probnorm(eta); - Probit link from the linear part to the expectation of the binary trait. You can choose the logit-link function rather than the probit link function using
mu=logistic(eta);
model $y \sim$ binary(mu); - Define the binary (also called the Bernoulli) density of data. You can use binary(mu) and the alias bern(mu) interchangeably.

Appendix C

Detailed explanations of the SAS code for the truncated Poisson data analysis

The SAS code is largely the same as the code in the Poisson data analysis. This appendix provides the explanations for the few extra statements.

if $y > 0$; - This statement in the data step selects observations of the SAS dataset when seeds are present. The subset of the sample is 204 in the fertility data.
fy=log(fact(y)); - Create a new variable fy=log(y!). The fact(y) function is the factorial of count y.
prior beta $\sim$ general(log(1)); - This statement defines a flat prior for variable β, i.e., $\pi(\beta) = 1$. Because the flat prior is not an intrinsic density, you must use the general function with the log density of the user defined prior as the argument. The log density is log(1).
prior sigmasqr&k~general(−log(signasqr&k)); - Define the Jeffreys’ prior for the kth variance component, $\pi(\sigma_k^2) = 1/\sigma_k^2$. The log prior density is $\log \pi(\sigma_k^2) = -\log \pi(\sigma_k^2)$.
$f = y*$ log(mu) − mu − fy; - Define logarithm of the Poisson density,

$$\begin{aligned}\log\left[f(y_j|\mu_j)\right] &= \log\left[\mu_j^{y_j}\exp(-\mu_j)/(y_j)!\right]\\ &= y_j\log(\mu_j) - \mu_j - \log\left[(y_j)!\right]\end{aligned}$$

$g = $ log(1 − exp(− mu)); - Define the logarithm of probability of $y > 0$, i.e., $\log[1 - \exp(-\mu_j)]$.
$f0 = f - g$; - This statement defines the logarithm of the truncated Poisson density,

$$\log\left[f_0(y_j|\mu_j)\right] = \log\left[f(y_j|\mu_j)\right] - \log\left[1 - \exp(-\mu_j)\right].$$

model $y \sim$ general($f0$); - The truncated Poisson density is a user defined density and thus it must be placed as an argument inside the general function.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)