A Bayesian test for Hardy-Weinberg equilibrium of bi-allelic X-chromosomal markers

Xavier Puig[†]

Josep Ginebra[†]

Jan Graffelman[†,‡]

[†]Department of Statistics and Operations Research

Universitat Politècnica de Catalunya

Barcelona, Spain.

[‡]Department of Biostatistics

University of Washington

Seattle WA, USA

- Corresponding author: Jan Graffelman, Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Avinguda Diagonal 647, 08028 Barcelona, Spain. email: jan.graffelman@upc.edu, tel: 00-34-934011739, fax: 00-34-934016575

Abstract

The X chromosome is a relatively large chromosome, harbouring a lot of genetic information. Much of the statistical analysis of X-chromosomal information is complicated by the fact that males only have one copy. Recently, frequentist statistical tests for Hardy-Weinberg equilibrium have been proposed specifically for dealing with markers on the X chromosome. Bayesian test procedures for Hardy-Weinberg equilibrium for the autosomes have been described, but Bayesian work on the X chromosome in this context is lacking. This paper gives the first Bayesian approach for testing Hardy-Weinberg equilibrium with bi-allelic markers at the X chromosome. Marginal and joint posterior distributions for the inbreeding coefficient in females and the male to female allele frequency ratio are computed, and used for statistical inference. The paper gives a detailed account of the proposed Bayesian test, and illustrates it with data from the 1000 Genomes project. In that implementation, a novel approach to tackle multiple testing from a Bayesian perspective through posterior predictive checks is used.

Key words: Hardy-Weinberg equilibrium; X-chromosome; Inbreeding coefficient; Bayesian model selection; Posterior distribution; Dirichlet prior;

1 Introduction

The number of genetic markers identified for the human genome has increased tremendously over the last decades. The 1000 genomes projects currently includes more than 88 million genetic variants (The 1000 Genomes Project Consortium, 2015). Most of the variants reside on the autosomes, which are ordered according to their size. The X chromosome is a large chromosome with a size of about 155Mb, and is almost as large as chromosome 7 (Hein et al., 2005), and estimated to contain about 5% of the genes in the human genome (Wise et al., 2013). Currently, approximately 3.5 million variants on the X chromosome have been reported. Much of the statistical analysis of the X chromosomal data is complicated by the fact that males have only one copy, whereas females have two. The pseudo-autosomal regions (Graves et al., 1998) of the X chromosome behave as autosomes, and for these regions autosomal statistical methodology applies.

A simple way to deal with X chromosomal data is to ignore males, and apply usual autosomal procedures to females only. This is what often has been done in studies of Hardy-Weinberg equilibrium, linkage disequilibrium, genetic association studies (Wise et al., 2013) and others. The Hardy-Weinberg law is a well-known elementary genetic principle typically explained in detail in genetic textbooks (Crow and Kimura, 1970; Li, 1976; Hartl, 1980; Hamilton, 2009). For a bi-allelic marker with alleles A and B with relative frequencies p and q, the law states that the genotype frequencies AA, AB and BB will reach the stable proportion $(p^2, 2pq, q^2)$ in one generation of random mating. From this point on, genotype and allele frequencies will remain unaltered through time, as long as disturbing forces like differential mortality, migration and others remain absent.

The dynamics of X-chromosomal markers is quite different. If male and female allele frequencies initially differ then it will take more than one generation before equilibrium is achieved. Because A males inherit their A allele from their mother, the male A allele fre-

quency always equals the female A allele frequency of the previous generation. Because females inherit one allele from each parent, the female A allele frequency is the mean of the male and female allele frequency of the previous generation. This "lagging and averaging" continues till the difference between male and female allele frequencies becomes vanishingly small. At that point, the female genotype frequencies will have stabilized as well, reaching the Hardy-Weinberg proportions. In each generation, the absolute difference between male and female allele frequencies is halved. If D_t represents the absolute difference in male and female allele frequency in generation t then we have $D_t = \left(\frac{1}{2}\right)^t D_1$, with D_1 the initial generation. In a worst case scenario with $D_1 = 1$, it will take 8 generations in total before the difference drops below 0.01. Figure 1 illustrates the faster attainment of HW equilibrium for smaller values of D_1.

Statistical tests for Hardy-Weinberg equilibrium should reflect the special characteristics of the X chromosome. In recent work, Graffelman and Weir (2016) have proposed chi-square, exact and permutation tests for Hardy-Weinberg (HW) equilibrium for markers at the X-chromosome that take both males and females into account. You et al. (2015) have developed a likelihood ratio test for X-chromosomal markers that also uses males and females. These frequentist procedures jointly test Hardy-Weinberg proportions for females and equality of allele frequencies in males and females.

There is a considerable number of contributions to the Bayesian testing for HW equilibrium of autosomal markers, starting with Pereira and Rogatko (1984) and Lindley (1988), and including Shoemaker et al. (1998), Ayres and Balding et al. (1998) and Wakefield (2009, 2010). Bayesian methods have also been used to deal with variants of unknown location, and classify them as autosomal or X-chromosomal under the assumption of HW equilibrium (Gautier, 2014). For testing autsomal variants for HW equilibrium, Ayres and Balding (1998) proposed an MCMC method to obtain the posterior distributions of inbreeding coefficients for markers with multiple alleles. Shoemaker et al. (1998) obtained

explicit expressions for the joint posteriors of various disequilibrium coefficients and allele frequencies in the biallelic case. Wakefield (2010) advocates the use of the Bayes Factor in Bayesian inference on HW equilibrium and addresses Bayesian testing in a genome-wide context. However, all these Bayesian studies address autosomal markers, and to date Bayesian procedures for X-chromosomal markers have apparently not been developed. This paper therefore first proposes Bayesian methods for a Hardy-Weinberg analysis of X-chromosomal markers that take both males and females into account, by using an extra parameter allowing for different allele frequencies in the sexes. We concentrate on the most commonly used single nucleotide polymorphisms (SNPs) and consider markers with multiple alleles, such as micro-satellites, beyond the scope of the current paper.

The structure of the paper is as follows. In Section 2 we provide some background, and establish notation. In Section 3 we develop a Bayesian approach to the problem of testing X-chromosomal markers for HW equilibrium, and we assess the method through simulation. Section 4 illustrates the use of the Bayesian approach with empirical data taken from the Japanese population of the 1000 Genomes project (The 1000 Genomes Project Consortium, 2010), both for single SNPs as well as sets of multiple X-chromosomal SNPs. The approach adopted in that implementation to deal with multiple testing through posterior predictive checks is novel. A discussion section completes the paper.

2 Background and notation

We consider a biallelic genetic polymorphism on the X chromosome with alleles A and B having allele frequencies p_{Am} and p_{Bm} in males and p_{Af} and p_{Bf} in females, with $p_{Am} + p_{Bm} = p_{Af} + p_{Bf} = 1$. There are five genotypes consisting of hemizygous males, with genotypes A and B, and diploid females, with genotypes AA, AB and BB. We denote the

observed genotype counts in males by n_{Am} and n_{Bm}, and in females by n_{AAf}, n_{ABf} and n_{BBf}. The total sample size is $n = n_m + n_f$, where $n_m = n_{Am} + n_{Bm}$ is the total number of males, and where $n_f = n_{AAf} + n_{ABf} + n_{BBf}$ is the total number of females.

The male A genotype (or allele) count, n_{Am}, is assumed to follow a Binomial(n_m, p_{Am}) distribution, and the vector of female genotype counts, $(n_{AAf}, n_{ABf}, n_{BBf})$, is assumed to follow a Multinomial$(n_f, (p_{AAf}, p_{ABf}, p_{BBf}))$ distribution, where $p_{AAf} + p_{ABf} + p_{BBf} = 1$.

2.1 Equilibrium in X-Chromosomal markers

For X-chromosomal markers it can take more generations to achieve equilibrium, depending on the initial difference in allele frequency between the sexes (Crow and Kimura, 1970). In fact, under disequilibrium, allele and genotype frequencies for the X chromosome will always be changing from generation to generation. All the frequencies considered next correspond to the current generation.

HW equilibrium holds for the SNPs of the X-chromosome when:

1. there is equality of male and female allele frequencies, $p_{Am} = p_{Af}$, and
2. the female genotype counts, $(n_{AAf}, n_{ABf}, n_{BBf})$, are multinomially distributed with HW proportions, $(p_{AAf} = p_{Af}^2, p_{ABf} = 2p_{Af}p_{Bf}, p_{BBf} = p_{Bf}^2)$, where $p_{Bf} = 1 - p_{Af}$.

When both these conditions hold in one generation then, under random mating, the allele frequencies in males and the genotype frequencies in females are constant from one generation to the next (see, e.g., Li, 1976, Zheng et al., 2007). In the case of X-chromosomal markers, disequilibrium can be present under three different scenarios.

In the first scenario, $p_{Am} = p_{Af}$ holds, but the female genotype proportions fail to match

the HW proportions, a case which is typically parametrized in terms of a female inbreeding coefficient, f, such that:

$$p_{AAf} = p_{Af}^2 + p_{Af}p_{Bf}f, \tag{2.1}$$

$$p_{ABf} = 2p_{Af}p_{Bf}(1-f), \tag{2.2}$$

$$p_{BBf} = p_{Bf}^2 + p_{Af}p_{Bf}f. \tag{2.3}$$

When $p_{Am} = p_{Af}$, a value of $f = 0$ corresponds to HW equilibrium, a positive f indicates a lack of female heterozygotes, and a negative f indicates an excess of female heterozygotes. Hence, for studying this first kind of disequilibrium on the X-chromosome we will use this female inbreeding coefficient, f, as a measure of the deviation of female genotype frequencies from HW proportions in the current generation, which can be posed as:

$$f = \frac{p_{AAf} - p_{Af}^2}{p_{Af}p_{Bf}}. \tag{2.4}$$

Note that the value of f can range between $-MAF/(1-MAF)$ and 1, where $MAF = \min(p_{Af}, 1-p_{Af})$. Under this first disequilibrium scenario and random mating, one will have HW equilibrium in the next generation, like for autosomal markers.

Under a second disequilibrium scenario, female genotype probabilities satisfy HW proportions and therefore $f = 0$, but the allele frequencies between males and females are different and therefore condition 1 does not hold, in which case we use as a measure for disequilibrium the ratio of male to female allele frequencies,

$$d = \frac{p_{Am}}{p_{Af}}. \tag{2.5}$$

Under this second disequilibrium scenario, with $d \neq 1$, allele frequencies of males and genotype frequencies of females converge to equilibrium only when the number of generations goes to infinity; Even though in this setting in the current generation f is 0, in the previous and in the following generations f is different from 0.

Under a third disequilibrium scenario, X-chromosomal markers might not be in equilibrium because both $f \neq 0$ as well as $d \neq 1$.

2.2 Models for equilibrium and for disequilibrium

In practice one will face either the HW equilibrium scenario, or one of three disequilibrium scenarios, which leads to the choice between four models. Under HW equilibrium, female genotype counts are Multinomial$(n_f, (p_{AAf} = p_{Af}^2, p_{ABf} = 2p_{Af}p_{Bf}, p_{BBf} = p_{Bf}^2))$ distributed, while the male A allele count follows a Binomial(n_m, p_{Am}) distribution with $p_{Am} = p_{Af}$. In this case, the value of p_{Af} determines the value of all the remaining probabilities. The model under HW equilibrium will be labeled as Model 0, (M_0).

In the first disequilibrium scenario described above, $f \neq 0$ and $d = 1$, and female genotype counts are Multinomial$(n_f, (p_{AAf}, p_{ABf}, p_{BBf}))$ distributed, while male A genotype counts are Binomial(n_m, p_{Am}) distributed with:

$$p_{Am} = p_{Af} = \frac{2p_{AAf} + p_{ABf}}{2}. \tag{2.6}$$

In this case the value of (p_{AAf}, p_{ABf}) determines the value of all the remaining probabilities. The model for this disequilibrium scenario is labeled as Model 1, (M_1).

In the second disequilibrium scenario described in section 2.1, the inbreeding coefficient for females, f, is equal to 0, but d is not equal to 1. In that case, female genotype counts are Multinomial$(n_f, (p_{AAf} = p_{Af}^2, p_{ABf} = 2p_{Af}p_{Bf}, p_{BBf} = p_{Bf}^2))$ distributed, while the male A genotype count is Binomial(n_m, p_{Am}) distributed with probability p_{Am} functionally unrelated to p_{Af}. In this case, the value of (p_{Af}, p_{Am}) determines the value of all the remaining probabilities. The model for this second disequilibrium scenario is labeled as Model 2, (M_2).

In the third and last disequilibrium scenario, $f \neq 0$ and $d \neq 1$. In that case, female genotype counts are Multinomial with unrestricted probabilities, as in Model 1, while the male A genotype count is Binomially distributed with probability p_{Am}, as in Model 2. In this case, the parameter space is the largest possible, and the model is labeled as Model 3, (M_3), or the saturated model.

3 Bayesian tests for X-chromosomal markers

In the frequentist approach to testing for HW equilibrium for X-chromosomal markers presented in Graffelman and Weir (2016), one chooses between HW equilibrium, (i.e. Model 0), and disequilibrium, (i.e. Models 1, 2 and 3), but it does not allow one to distinguish between the three different disequilibrium scenarios. In the frequentist approach, additional statistical tests for equality of allele frequencies and/or HW proportions in females would be needed to finally pinpoint the scenario.

Instead, in the Bayesian setting it is more natural to test for HW equilibrium by choosing one scenario among the four alternative scenarios described above, which is equivalent to selecting one model among M_0, M_1, M_2 and M_3. That is done by choosing a prior distribution for the parameters of the models that captures what one knows about them before observing the data, and a prior distribution on the model space, and then computing the posterior probability of each one of the four models (scenarios). Then, one selects the model (scenario) with largest posterior probability.

3.1 Choice of a prior distribution

Different parametrizations allow for different ways of capturing what one knows about the parameters of the model in terms of a prior distribution for them. Here we will adopt the parametrization of Models 0, 1, 2, 3 in terms of male and female genotype frequencies, because that allows for a choice of priors that leads to simple expressions for the posterior probabilities of the four models considered, and because they are the most convenient ones when one has little information.

Under the HW equilibrium scenario, leading to Model 0, male and female allele frequencies, p_{Am} and p_{Af}, are equal, and they will be assumed to be Beta$(b_{1,0}, b_{2,0})$ distributed, where the second subindex, 0, refers to M_0. Under this scenario, this prior distribution univocally determines the prior distribution of all female genotype frequencies.

Under the first disequilibrium scenario, leading to Model 1, the female genotype frequencies, $(p_{AAf}, p_{ABf}, p_{BBf})$, are assumed to be Dirichlet$(a_{1,1,f}, a_{2,1,f}, a_{3,1,f})$ distributed, where the second subindex, 1, refers to M_1. The distribution of the female genotype frequencies determines the distribution of the female and male allele frequencies.

Under the second disequilibrium scenario, leading to Model 2, male and female allele frequencies are assumed to be independently distributed as a Beta$(b_{1,2,m}, b_{2,2,m})$ and Beta$(b_{1,2,f}, b_{2,2,f})$ respectively, and that determines the distribution of the female genotype frequencies. Finally, under the last disequilibrium scenario, leading to Model 3, female genotype frequencies are assumed to be Dirichlet$(a_{1,3,f}, a_{2,3,f}, a_{3,3,f})$ distributed, independent of the male allele frequency, which is assumed to be Beta$(b_{1,3,m}, b_{2,3,m})$ distributed.

Depending on the values chosen for $(a_{1,i,f}, a_{2,i,f}, a_{3,i,f})$, the Dirichlet$(a_{1,i,f}, a_{2,i,f}, a_{3,i,f})$ distribution will be more or less informative, and it will capture different information about female genotype frequencies. In particular, its expected value is $(a_{1,i,f}, a_{2,i,f}, a_{3,i,f})/(\sum_{j=1}^{3} a_{j,i,f})$,

and one can choose the $a_{j,i,f}$'s to reflect the fact that one expects some genotypes to have larger probabilities than others. Also, the larger $\sum_{j=1}^{3} a_{j,i,f}$, the smaller the variances of the components of the Dirichlet random variable, and the more informative that prior distribution. When one is not willing to use subjective information about the female genotype frequencies, Berger et al. (2015) recommend using a Dirichlet with $a_{1,i,f} = a_{2,i,f} = a_{3,i,f} = 1/3$, which is also recommended by Bernardo and Tomazella (2010) as a good approximation to a prior distribution tailored for a reference analysis of HW equilibrium. We will use this reference prior, which is like assuming an effective sample size of only one to start with (see, e.g., Morita et al., 2008). Given that the actual sample sizes in our setting will typically be a lot larger than that, the impact of this prior on the posterior distribution for female genotype frequencies will be negligible.

An analogous argument can be made for choosing the parameters of the Beta$(b_{1,i}, b_{2,i})$ to model the prior information about allele frequencies. In that case, in the absence of subjective information one often chooses Beta$(b_{1,i}, b_{2,i})$ with $b_{1,i} = b_{2,i} = 1/2$, which corresponds to a relatively uninformative prior that assumes an effective sample size of only one to start with. Moreover, this prior captures the fact that low MAF markers are more frequent.

An alternative way of eliciting prior information under Models 1 and 3 is to choose a specific distribution for the inbreeding coefficient, f, and for the female allele frequency, p_{Af}, instead of resorting to the Dirichlet distribution for the genotype frequencies. However, this complicates the computation of the posterior probabilities, and it does not make much difference when carrying out a reference analysis that uses little prior information. The inbreeding coefficient is related to the female genotype frequencies through:

$$f = \frac{p_{AAf} - (p_{AAf} + \frac{1}{2}p_{ABf})^2}{(p_{AAf} + \frac{1}{2}p_{ABf})(1 - p_{AAf} - \frac{1}{2}p_{ABf})}, \tag{3.1}$$

and one can explore the prior distribution of f that is induced by assuming a Dirichlet distribution on $(p_{AAf}, p_{ABf}, p_{BBf})$. When one does that for our reference choice, with

$a_{1,i,f} = a_{2,i,f} = a_{3,i,f} = 1/3$, one finds that the prior distribution for f is not symmetric on its support. That prior is in fact trimodal, with two modes at the two extremes of the range of values taken by f, and a third mode at 0, which are features that one considers desirable for a reference prior for a parameter such as f, with finite range and a null hypothesis at 0. Furthermore, under our choice of parameters for the Dirichlet prior, one can check through Monte Carlo simulation that the probability that f is larger than 0 is .548.

Instead, when one assumes a Dirichlet$(a_{1,i,f}, a_{2,i,f}, a_{3,i,f})$ distribution with $a_{1,i,f} = a_{2,i,f} = a_{3,i,f} = 1$ as the prior distribution for the female genotype frequencies, which corresponds to assuming a uniform distribution on them, one finds that the prior distribution for f concentrates on values larger than 0, with a prior probability that f is larger than 0 equal to .667. The problem of this upward bias introduced when using a uniform distribution has already been reported by Foll and Gaggiotti (2008). Another shortcoming of assuming a uniform prior for $(p_{AAf}, p_{ABf}, p_{BBf})$ is that the prior induced on the female allele probability through $p_{Af} = (2p_{AAf} + p_{ABf})/2$ becomes strongly unimodal with mode at .5. Instead, our choice of $a_{1,i,f} = a_{2,i,f} = a_{3,i,f} = 1/3$ leads to a prior distribution for p_{Af} that is a lot closer to the Beta$(.5, .5)$ that is assumed for p_{Am}.

Note though that either one of these two choices of values for $(a_{1,i,f}, a_{2,i,f}, a_{3,i,f})$ leads to posterior distributions that are very similar, because they are both a lot less informative than the data that one typically obtains in these settings.

Alternative ways of choosing prior distributions for HW equilibrium under the usual autosomal data can be found in Lindley (1988), Shoemaker et al. (1998), Consonni et al. (2008) and Wakefield (2010). All their proposals could be adapted to our X-chromosomal marker setting, but if one chose these priors to have a small effective sample size, they would make a small difference at a considerable extra computational cost, because they do not lead to closed form expressions for the posterior probabilities described next.

3.2 Bayesian model selection

The Bayesian way to select a model is through the posterior probability of each model, $P(M_i|y)$, which is the probability that the M_i model is the one generating the data, $y = (n_{AAf}, n_{ABf}, n_{BBf}, n_{Am}, n_{Bm})$, assessed after the data has been observed. It can be computed by using Bayes theorem:

$$P(M_i|y) = \frac{P(M_i)P(y|M_i)}{\sum_{j=0}^{3} P(M_j)P(y|M_j)}, \qquad \text{for} \qquad i = 0, 1, 2, 3 \tag{3.2}$$

where $P(M_i)$ is the prior probability assigned to M_i, (i.e. the probability that this model is correct, assessed before the data is available), and where $P(y|M_i)$ is the marginal likelihood of M_i. If all models were considered equally likely a priori, the way it will be assumed in Section 4, the larger $P(y|M_i)$, the more attractive M_i will be.

Most often, computing $P(y|M_i)$ exactly is too complicated, and the marginal likelihoods need to be estimated through the MCMC simulations used to update the model. In our binomial/multinomial setting with Beta/Dirichlet priors though, there are closed form expressions for $P(y|M_i)$ which allow one to either compute these marginal likelihoods exactly, in the case of Models 0, 2 and 3, or to evaluate them numerically in the case of Model 1. The expressions for the marginal likelihoods, $P(y|M_i)$, under our choice of prior distribution can be found in the Appendix 1; They allow one to compute the posterior probabilities on the model space, $P(M_i|y)$, exactly through (3.2).

To assess the strength of evidence in favor or against a given model, M_i, one sometimes resorts to the corresponding Bayes factor, BF_i, which is the ratio of the posterior odds and the prior odds for that model. When all four models are considered equally likely a priori, $BF_i = 3P(M_i|y)/(1-P(M_i|y))$. One usually considers that a $\log_{10}(BF_i)$ that takes a value between .5 and 1 indicates that the strength of evidence in favor of M_i is substantial, when its value is between 1 and 1.5 it is strong, when it is between 1.5 and 2 it is very strong,

and when it is larger than 2 it considers the evidence in favor of M_i to be decisive.

3.3 Simulation assessment of the Bayesian test

To assess the performance of this Bayesian test for HW equilibrium, here it is used under a very wide set of known scenarios through an extensive simulation study. In particular, the test is tried on SNPs from populations with inbreeding coefficients, f, taking values in its whole range, and with a ratio of male to female allele frequencies, d, ranging between 0.5 and 2. In total, we have considered 625 different pairs of values for (f, d), and for each pair we have checked the performance of the test on populations with $p_{Af} = 0.2$ and 0.4 assuming samples with $n_f = n_m = 500$ and with $n_f = n_m = 2000$.

For each one of the set of 2500 values of (f, d, p_{Af}, n_f) considered we have simulated 1000 independent SNPs with sample size $n = n_f + n_m$ from a population with the corresponding values of (f, d, p_{Af}), and we have computed $P(M_i|y)$ for $i = 0, 1, 2, 3$ and for each one of the samples.

Figure 2 presents the contour plots for the average of all the values of $P(M_i|y)$ obtained, as a function of (f, d) for the four combinations of (p_{Af}, n_f) considered. This average estimates the expected value of $P(M_i|y)$ for each given (f, d, p_{Af}, n_f). As desirable, the expected value of $P(M_i|y)$ peaks on the region of the (f, d) space where the corresponding M_i model holds true. One also observes that the larger the sample size n, and/or the larger p_{Af}, the more peaked the expected value of $P(M_i|y)$ is as a function of (f, d), and hence the better does this Bayesian test work.

4 Examples

To present applications of the Bayesian approach to testing for HW equilibrium advocated in this paper, we analyze individual markers (subsection 4.1) and groups of markers (subsection 4.2) of the Japanese population of the 1000 Genomes project, consisting of $n_m = 56$ males and $n_f = 48$ females. We also explain how one can take into account the multiple testing effect through posterior predictive checks when assessing the HW equilibrium hypothesis based on the simultaneous analysis of a large number of SNPs (subsection 4.3).

4.1 Test on four individual SNPs

In order to compare the Bayesian test proposed here for HW equilibrium at biallelic genetic markers on the X-chromosome with the tests proposed in the context of a frequentist approach, we report the posterior probabilities of the four possible scenarios together with the p-values of the exact tests for four example SNPs in Table 1. Exact tests were performed with and without the data on males using the methods proposed by Graffelman and Weir (2016). The posterior probabilities are computed through (3.2), assuming equal prior probabilities for the four models, and hence $P(M_i) = 1/4$, and using the expressions for the marginal likelihoods, $P(y|M_i)$ in the Appendix 1 with $a_{j,i,f} = 1/3$ for the Dirichlet prior and $b_{j,i} = 1/2$ for the beta priors. Given that each one of these priors corresponds to an effective sample size of only one and data involves a sample size of $n = 104$, the role played by the prior distribution is negligible. Sample sizes will most often be larger than in this example, and hence in practice the choice of a prior will most often be even less relevant.

The first marker in Table 1, rs13440889, has a posterior probability of .748 of being in HW equilibrium, and hence one rejects the three disequilibrium scenarios, with posterior

probabilities of .126 or smaller. The corresponding Bayes factor indicates that the evidence in favor of being in HW equilibrium here is substantial. This is consistent with the non-significant exact test for HWE (p = .954). The second marker in Table 1, has a posterior probability of only .072 of being in HW equilibrium, but it has instead a posterior probability of .803 of being in the first disequilibrium scenario, with $d = 1$ and $f \neq 0$, and hence one settles with M_1 for that marker. Here the Bayes factor indicates that the evidence in favor of M_1 is strong. In this case, choosing M_1 is in agreement with the frequentist exact test rejecting HW proportions in females (p = .004).

For the third marker in Table 1, HW equilibrium is also rejected, because it has a posterior probability of only .092, and one settles with the second disequilibrium scenario, with $d \neq 1$ and with $f = 0$. Note that in this case the frequentist tests do reject HWE overall, but do not reject HW proportions for females (p = .760). For the last marker in Table 1, the most probable scenario is clearly the third disequilibrium scenario, with $d \neq 1$ and $f \neq 0$. Here BF_3 indicates that the evidence in favor of M_3 is decisive. The frequentist tests reject equilibrium (p < .0005), but the difference in allele frequencies goes unnoticed.

In Figure 3 one has the set of marginal posterior distributions for the marker in the second row in Table 1, SNP rs2301322. These marginal posteriors are computed assuming the full Model 3, in the way described in Appendix 2. The first row presents the marginal posterior for female genotype frequencies, the second row presents the marginal posterior for male and female allele frequencies and for their ratio, while the third row presents the marginal posterior for the inbreeding coefficient as well as the joint posterior for allele frequencies and for (f, d).

Figure 3 also presents 90% highest posterior density (hpd) credible intervals/regions for all these parameter values or pairs of parameter values. The marginal posterior for f in Figure 3, for example, places almost all its probability mass away from $f = 0$, with the 90% hpd

SNP	Genotypes					Posterior probabilities and BF				P-values	
	n_{AAf}	n_{ABf}	n_{BBf}	n_{Am}	n_{Bm}	$P(M_0\|y)$ $log_{10}BF_0$	$P(M_1\|y)$ $log_{10}BF_1$	$P(M_2\|y)$ $log_{10}BF_2$	$P(M_3\|y)$ $log_{10}BF_3$	all	female
rs13440889	26	19	3	43	13	0.748 0.949	0.126 -0.363	0.107 -0.442	0.019 -1.246	0.954	1.000
rs2301322	33	9	6	44	12	0.072 -0.632	0.803 1.087	0.010 -1.535	0.115 -0.408	0.009	0.004
rs2356583	6	25	17	35	21	0.092 -0.516	0.016 -1.300	0.744 0.941	0.147 -0.286	0.015	0.760
rs201728945	3	45	0	44	12	0.000 -10.696	0.001 -2.741	0.000 -9.353	0.999 3.695	0.000	0.000

Table 1: Genotype counts, posterior probabilities and the base 10 logarithm of the Bayes factor of the four scenarios (M_0: HW equilibrium, M_1: $f \neq 0$, M_2: $d \neq 1$, M_3: $\{f \neq 0$ and $d \neq 1\}$), and exact p-values for the all-individual and for the females-only test, for four single-nucleotide polymorphisms from a sample of the Japanese population study with $n_m = 56$ and $n_f = 48$.

posterior credible interval being $(0.199, 0.683)$. Instead, the marginal posterior for d places $d = 1$ well inside its 90% hpd posterior credible interval, which is $(0.853, 1.172)$. The results clearly show that females are out of HW proportions, but that equality of male and female allele frequencies is a tenable supposition.

Note that different from confidence regions, Bayesian credible regions are statements about the probability that the actual parameter value for that given SNP falls in a given region, and not the probability that the region captures the true parameter value under repeated use of these regions on different samples.

Figure 4 presents the marginal posterior distributions for (f, d) for the four SNPs in Table 1, together with its 90% hpd posterior credible region. The fact that, for example, for the SNP rs13440889, the $(0, 1)$ point falls well inside the 90% posterior credible region is a clear

indication that in that case HW equilibrium holds. In the other three examples the $(0, 1)$ point falls outside the corresponding 90% credible region in three different ways, which are representative of the three different reasons through which equilibrium might be broken.

4.2 Simultaneous analysis of multiple X-chromosomal SNPs

In this section we illustrate the Bayesian approach to testing for HW equilibrium of X-chromosomal markers by carrying out the Bayesian test based on the simultaneous analysis of a large set of SNPs selected from the Japanese population of the 1000 Genome project. The 1000 Genomes project provides genotype information for approximately 3.5 million variants on the X chromosome. SNPs without rs identifier, SNPs in the pseudo-autosomal regions, and SNPs with missing values were excluded. X-chromosomal SNPs were linkage disequilibrium pruned with Plink (Purcell et al., 2007) using the independent pairwise option with a sliding window of 50 SNPs and a threshold of $R^2 = 0.50$ using Plink instruction `plink --bfile JPTChrX --indep-pairwise 50 5 0.50 --ld-xchr 1`). The SNPs with small MAF have not been filtered out. This leaves a sample of 162225 SNPs from the whole X chromosome, that is the one that will be used in this subsection.

Figure 5 presents the model with the largest posterior probability for each one of these SNPs, presented in the order in which these SNPs appear on the X-chromosome. The white band without SNPs between 58.1 MB and 63.0 Mb corresponds to the centromere. The presence of consecutive sequences of markers being systematically classified to the same disequilibrium scenario, or to one of the three disequilibrium scenarios, might be an indication of quality control problems in the SNP measurements, or might arise if the PAR region is erroneously included in the analysis. Too few SNPs being classified as being in HW equilibrium would also be an indication of either a problem in the measurements or of the fact that the population under scrutiny is actually in disequilibrium.

Model 0, representing HW equilibrium, is the one with the largest posterior probability in 95.27% of all the 162225 SNPs considered, Model 1 is the one with the largest probability in 1.89% of the cases, Model 2 is the one with the largest probability in 2.13% of the cases, and Model 3 is the one with the largest probability in 0.71% of the cases. It is known that for SNPs with low MAF, which are abundant in this set of 162225 SNPs, power to detect disequilibrium is low. When one filters out the SNPs with $MAF < 0.05$, one is left with only 52008 SNPs, and the proportion classified as being in equilibrium falls down to 89.08%.

The next subsection illustrates how one can assess whether the overall proportions obtained for all 162225 SNPs are compatible with the HW equilibrium model, M_0, holding true, in a way that takes into account the multiple testing effect involved.

4.3 Multiple testing and the assessment of HW equilibrium

Carrying out the test for HW equilibrium based on the simultaneous analysis of multiple SNPs involves dealing with the multiple testing effect, which requires one to account for the experiment-wise error rate. In the Bayesian context, one approach to that problem is through the use of the false discovery rate (FDR) and q-values, as described in Storey (2002, 2003), Muller et al. (2006), de Villemereuil et al. (2014) and de Villemereuil and Gaggiotti (2015).

Instead of using the FDR, here a novel approach to address multiple testing in the Bayesian setting is used. The alternative method uses posterior predictive checks to assess whether the proportion of SNPs in the sample of 162225 of Subsection 4.2 classified to each one of the four different scenarios is consistent with the proportions that would be obtained if the HW equilibrium was actually in place for the population. To estimate the proportions classified

under each scenario in a population of SNPs actually in HW equilibrium, simulation from the posterior predictive distribution under M_0 is used. For a description of the use of posterior predictive checks as a tool to validate models in general, see Chapter 6 of Gelman et al. (2014), or Puig and Ginebra (2014).

To do this simulation exercise, one needs to resort to a sample of SNPs that is smaller than the one used in Subsection 4.2, because we need to assume approximate independence between SNPs and because, at this point, the simulation exercise that would need to be done with the larger set of SNPs would take too long. That is why a random subsample of only 1622 SNPs is obtained from the 162225 SNPs used in Subsection 4.2 to carry out the posterior predictive checks.

It turns that for the subsample with only 1% of all the SNPs previously used, Model 0 is the one with the largest posterior probability in 95.25% of the SNPs, Model 1 is the one with the largest probability in 2.03% of the cases, Model 2 is the one with the largest probability in 2.03% of the cases, and Model 3 is the one with the largest probability in 0.68% of the cases. The second panel in Figure 5 presents the model with the largest probability for each one of these 1622 SNPs.

To assess whether these observed proportions of SNPs being classified as following each one of the models are compatible with the assumption that HW equilibrium is in place, we estimate the posterior predictive distribution and the posterior predictive credible intervals for these four proportions, assuming that the HW equilibrium holds and that the SNPs are independent. This last assumption will be satisfied due to the way in which the smaller subset of only 1622 SNPs was selected from the whole set of SNPs of the X-chromosome.

The posterior predictive distribution can be estimated by repeatedly simulating 1622×5 tables of "data like the one from the Japanese study" used to test for HW equilibrium, by using the posterior predictive distribution of the data under Model 0. For each one of the

simulated tables, one then classifies the 1622 simulated SNPs to one of the four scenarios, based on their $P(M_i|y)$ and finds the proportion of SNPs classified into each scenario.

Each one of the tables can be simulated from the posterior predictive distribution by:

1. Simulating 1622 values for p_{Af}, one value of each row of the table, using its posterior distribution under Model 0 which, assuming Beta$(b_{1,0}, b_{2,0})$ to be the prior, is:

$$\pi(p_{Af}|y) \sim \text{Beta}(b_{1,0} + n_{Am} + 2n_{AAf} + n_{ABf}, b_{2,0} + n_{Bm} + 2n_{BBf} + n_{ABf}), \quad (4.1)$$

   which is also the posterior for p_{Am} because under M_0, $p_{Am} = p_{Af}$.

2. For each value of p_{Am} one simulates the n_{Am} for that row from a binomial(n_m, p_{Am}), one computes $n_{Bm} = n_m - n_{Am}$, and one simulates $(n_{AAf}, n_{ABf}, n_{BBf})$ for that row from a Multinomial$(n_f, (p_{Af}^2, 2p_{Af}(1-p_{Af}), (1-p_{Af})^2))$.

3. For each row of each table one computes $P(M_i|y)$ for $i = 0, 1, 2, 3$, and for each simulated table one obtains the proportions of SNPs classified as following each one of the four M_i's based on the largest $P(M_i|y)$ for that row.

By repeating this exercise as many times as tables of data one intends to simulate, one obtains the posterior predictive distribution for the proportions of SNPs being classified as M_i for $i = 0, 1, 2, 3$, conditioned on the HW equilibrium model, M_0, being the correct one. The rows of the new tables are simulated to be independent, which is a realistic assumption when one is analyzing a subset of approximately independent SNPs the way it is done here.

We have carried out this simulation exercise for the subset of 1622 SNPs selected from the Japanese population study by simulating 1000 tables from its posterior predictive distribution. It turns that if the HW equilibrium is in place, the 90% central posterior predictive credible interval for the proportion of SNPs classified as following M_0 because their $P(M_0|y)$

is the largest is (94.6,96.4), the 90% credible interval for the proportion of SNPs classified as following M_1 because $P(M_1|y)$ is the largest is (1.4,2.3), the one for the proportion of SNPs classified as following M_2 is (1.8,3.0), and the one for the proportion of SNPs classified as following M_3 is (0.1,0.5).

Note that the proportion of SNPs being classified into each one of these four models for the subset of markers from the Japanese population Genome project, which are 95.25%, 2.03%, 2.03% and 0.68% for M_0, M_1, M_2 and M_3, fall either well within these four posterior predictive credible intervals, or very close to it in the case of M_3. The fact that the observed percentages fall within the posterior predictive intervals generated under the HWE assumption, suggests that the X-chromosomal markers without missing values of the LD-pruned database are in equilibrium, with only a slight excess of markers in scenario M_3. By using posterior predictive checks involving all 1622 SNPs at once, instead of doing it one SNP at a time, one already takes into account the experiment-wise error rate, and one does not have to correct for the fact that one carries out multiple tests.

When one does the same exercise on data from populations that are not in HW equilibrium, the proportion of SNPs that are classified as following M_0 falls, and some of the other three proportions increase, and they would fall outside of the posterior predictive intervals for these four proportions obtained assuming that the HW equilibrium model was in place. Given that the sample size here is a lot larger than the effective sample size assumed by the priors, carrying out a sensitivity analysis that considers alternative priors of similar effective sample size leads to results which are almost identical to the ones reported here.

5 Discussion

We have developed a Bayesian method for inference on Hardy-Weinberg equilibrium for bi-allelic markers at the X-chromosome. Disequilibrium at the X chromosome may be due to a difference in allele frequencies between the sexes, or to females not corresponding to HW proportions, or both these factors simultaneously. By computing the posterior probability for each scenario, geneticists can immediately infer the most likely scenario. A similar approach can also be used for the Bayesian analysis of autosomal variants.

The X-chromosomal exact test chooses between HW equilibrium, (i.e. Model 0), and dis-equilibrium, (i.e. Models 1, 2 and 3). In order to precisely determine the disequilibrium scenario with a frequentist approach, several statistical tests are necessary: an exact test with and without males and eventually an exact test for equality of male and female allele frequencies. Instead, by assigning a posterior probability to each one of the four scenarios, with the four probabilities adding up to one, our Bayesian approach provides a simple way of selecting the most probable scenario in the light of the data.

One of the advantages of the Bayesian approach to HW equilibrium testing of X-chromosomal markers is that, on top of yielding posterior probabilities for each one of the four scenarios, it also provides the posterior distribution of the parameters of interest. In Appendix 2 one can find details on that distribution.

Among all the marginal posterior distributions, the one for (f, d) is particularly useful because it helps one assess the degree of departure from HW equilibrium beyond computing the corresponding four posterior probabilities.

For our Bayesian analysis, we have found it convenient to parametrize disequilibrium by using the inbreeding coefficient and the ratio of male to female allele frequencies, using a

Dirichlet prior on the genotype frequencies. Alternatively, other disequilibrium measures with priors specified directly on the disequilibrium measures might also be considered.

One side contribution of this manuscript is the suggestion to use posterior predictive checks to deal with multiple testing in the Bayesian framework, as described in subsection 4.3.

From a computational point of view, the chi-square test for HWE of X-chromosomal markers is very fast, and it is feasible to do this for a complete X chromosome with 3.5 million markers. An exact test is computationally more demanding due to the presence of factorial calculations and enumeration of possible outcomes. The Bayesian procedures outlined in this paper do not require a MCMC implementation as it is usual in most Bayesian applications these days, and that simplifies the computation a lot. If the integration required for the computation of the posterior probability of M_1 is carried out efficiently, there should not be any problem in using the proposed method for a whole X chromosome.

Further computational savings could be attained by using the fact that many of the 3.5 million markers on the X chromosome are rare variants with a low minor allele frequency, and therefore the set of genotype counts will be identical for many SNPs. For markers with identical counts, the HW tests only have to be computed once.

6 Software

The Bayesian X-chromosomal procedures described in this paper have been programmed in R (R Core Team, 2014) by Xavi Puig, and are made available in version 1.5.7 of the HardyWeinberg package (Graffelman, 2015).

References

Ayres, K.L., Balding, D.J. (1998). Measuring departures from Hardy-Weinberg: a Markov chain monte carlo method for estimating the inbreeding coefficient. Heredity, 80: 769-777.

Berger, J.O., Bernardo, J.M., Sun, D. (2015). Overall objective priors. Bayesian Analysis, 10: 189-221.

Bernardo, J., Tomazella, V. (2010). Bayesian reference analysis of the Hardy-Weinberg equilibrium. In M.H.Chen, D.K.Dey, P.Muller, D.Sun and K.Ye (eds.), Frontiers of Statistical Decision Making and Bayesian Analysis, In Honor of James O. Berger, p. 31-43. Springer Verlag.

Consonni, G., Gutierrez-Pena, E. and Veronese, P. (2008). Compatible priors for Bayesian model comparison with an application to the Hardy-Weinberg equilibrium model. Test, 17, 585-605.

Crow, J.F., Kimura, M. (1970). An Introduction to Population Genetics Theory. Harper & Row Publishers.

Foll, M., Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics, 180: 977-993.

Gang, Z., Joo, J., Zhang, C., Geller, N.L. (2007). Testing association for markers on the X chromosome. Genetic Epidemiology, 31: 834-843.

Gautier, M. (2014). Using genotyping data to assign markers to their chromosome type and to infer the sex of individuals: a Bayesian model-based classifier. Molecular Ecology Resources, 14, 1141-1159.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B. (2014). Bayesian Data Analysis, 3rd Ed. Chapman and Hall.

Graffelman, J. (2015). Exploring diallelic genetic markers: the Hardy-Weinberg package. Journal of Statistical Software, 64, 1-23.

Graffelman, J., Morales-Camarena, J. (2008). Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. Human Heredity, 65: 77-84.

Graffelman, J., Weir, B.S. (2016). Testing for Hardy-Weinberg equilibrium at biallelic genetic markers on the X chromosome. Heredity, 116(6): 558–568.

Graves, J.A., Wakefield, M.J. and Toder, R. (1998) The origin and evolution of the pseudoautosomal regions of human sex chromosomes. Human molecular genetics, 7(13): 1991–1996.

Hamilton, M.B. (2009). Population Genetics. Wiley-Blackwell.

Hartl, D.L. (1980). Principles of Population Genetics. Sinauer Associates, Sunderland, Massachusetts.

Hein, J., Schierup, M.K. and Wiuf, C. (2005). Gene Genealogies, Variation and Evolution. Oxford University Press.

Li, C.C. (1976). The First Course in Population Genetics. California, The Boxwood Press.

Lindley, D. (1988). Statistical inference concerning Hardy-Weinberg equilibrium. In J.Bernardo, M.DeGroot, D.Lindley and A.Smith (eds.), Bayesian Statistics 3, 307-320. Oxford: Oxford University Press.

Montoya-Delgado, L., Irony, T., Pereira, C., Whittle, M. (2001). An unconditional exact test for the Hardy-Weinberg equilibrium law: Sample-space ordering using the Bayes factor. Genetics, 158, 875-883.

Morita, S., Thall, P.F., Muller, P. (2008). Determining the effective sample size of a parametric prior. Biometrics, 64: 595-602.

Muller, P., Parmigiani, P. and Rice, K. (2006). FDR and Bayesian multiple comparison rules. Johns Hopkins University, Department of Statistics Working Papers 115.

Pereira, C., Rogatko, A. (1984). The Hardy-Weinberg equilibrium under a Bayesian perspective. Revista Brasileira de Genetica, 4, 689-707.

Puig, X., Ginebra, J. (2014). A Bayesian cluster analysis of election results. Journal of Applied Statistics, 41, 73-94.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. and Sham, P.C. (2007). PLINK: A Toolset for Whole-Genome Association and Population-Based Linkage Analysis. American Journal of Human Genetics, 81(3) 559–575. http://pngu.mgh.harvard.edu/purcell/plink/.

Shoemaker, J., Painter, I. and Weir, B. (1998). A Bayesian characterization of Hardy-Weinberg disequilibrium. Genetics, 149: 2079-2088.

Storey, J.D. (2002). A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B, 64: 479-498.

Storey, J.D. (2003). The positive false discovery rate. A Bayesian interpretation and the q-value. Annals of Statistics, 31: 2013-2035.

The 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M. et al. (2010). A map of human genome variation from population-scale sequencing. Nature, 467: 1061-1073.

The 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M. et al. (2015). A global reference for human genetic variation. Nature, 526:68-74.

de Villemereuil, P., Frichot, E., Bazin, E., François, O., Gaggiotti, O.E. (2014). Genome scan methods against more complex models: when and how much should we trust them? Molecular Ecology, 23: 2006-2019.

de Villemereuil, P., Gaggiotti, O.E. (2015). A new F_{ST}-based method to uncover local adaptation using environmental variables. Methods in Ecology and Evolution, 6, 1248-1258.

Wakefield, J. (2009). Bayes factors for genome-wide association studies: Comparison with p-values. Genetic Epidemiology, 33, 79-86.

Wakefield, J. (2010). Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics, 66: 257-265.

Wise, A.L., Gyi, L., Manolio, T.A. (2013) eXclusion: toward integrating the X chromosome in genome-wide association analyses. American Journal of Human Genetics, 92: 643-647.

You, X.P., Zou, Q.L., Li, J.L., Zhou, J.Y. (2015). Likelihood ratio test for excess homozygosity at marker loci on X chromosome. PLOS One, 10(12): e0145032.

Zheng, G., Joo, J., Zhang, C., Geller, N. L. (2007) Testing association for markers on the X chromosome. Genetic Epidemiology, 31: 834–843. 10.1002/gepi.20244.

Appendix 1; Marginal likelihoods

Here we present the marginal likelihoods, $P(y|M_i)$ for $i = 0, \ldots, 3$, needed to compute the posterior probabilities, $P(M_i|y)$, through (3.2). The priors assumed are the ones described in Section 3.1, and $y = (n_{AAf}, n_{ABf}, n_{BBf}, n_{Am}, n_{Bm})$.

The marginal likelihood under Model 0, under HW equilibrium, is:

$$P(y|M_0) = \frac{n_f!}{n_{AAf}!n_{ABf}!n_{BBf}!}\frac{n_m!}{n_{Am}!n_{Bm}!}\frac{\Gamma(\sum_{j=1}^{2} b_{j,0})}{\prod_{j=1}^{2}\Gamma(b_{j,0})}2^{n_{ABf}}\times$$

$$\frac{\Gamma(b_{1,0}+2n_{AAf}+n_{ABf}+n_{Am})\Gamma(b_{2,0}+2n_{BBf}+n_{ABf}+n_{Bm})}{\Gamma(\sum_{j=1}^{2} b_{j,0}+2n_f+n_m)}. \quad (6.1)$$

The marginal likelihood under Model 1, with $d = 1$ and $f \neq 0$, can be computed through:

$$P(y|M_1) = \frac{n_f!}{n_{AAf}!n_{ABf}!n_{BBf}!}\frac{n_m!}{n_{Am}!n_{Bm}!}\frac{\Gamma(\sum_{j=1}^{3} a_{j,1,f})}{\prod_{j=1}^{3}\Gamma(a_{j,1,f})}\int_0^1\int_0^{1-p_{ABf}} p_{AAf}^{a_{1,1,f}+n_{AAf}-1}\times$$

$$p_{ABf}^{a_{2,1,f}+n_{ABf}-1}(1-p_{AAf}-p_{ABf})^{a_{3,1,f}+n_{BBf}-1}\left(\frac{2p_{AAf}+p_{ABf}}{2}\right)^{n_{Am}}\left(1-\frac{2p_{AAf}+p_{ABf}}{2}\right)^{n_{Bm}}dp_{AAf}dp_{ABf}. \quad (6.2)$$

The marginal likelihood under Model 2, with $d \neq 1$ and $f = 0$, is:

$$P(y|M_2) = \frac{n_f!}{n_{AAf}!n_{ABf}!n_{BBf}!}\frac{n_m!}{n_{Am}!n_{Bm}!}\frac{\Gamma(\sum_{j=1}^{2} b_{j,2,f})}{\prod_{j=1}^{2}\Gamma(b_{j,2,f})}\frac{\Gamma(\sum_{j=1}^{2} b_{j,2,m})}{\prod_{j=1}^{2}\Gamma(b_{j,2,m})}2^{n_{ABf}}\times$$

$$\frac{\Gamma(b_{1,2,f}+2n_{AAf}+n_{ABf})\Gamma(b_{2,2,f}+2n_{BBf}+n_{ABf})}{\Gamma(\sum_{j=1}^{2} b_{j,2,f}+2n_f)}\frac{\Gamma(b_{1,2,m}+n_{Am})\Gamma(b_{2,2,m}+n_{Bm})}{\Gamma(\sum_{j=1}^{2} b_{j,2,m}+n_m)}. \quad (6.3)$$

Finally, the marginal likelihood under the saturated Model 3 is:

$$P(y|M_3) = \frac{n_f!}{n_{AAf}!n_{ABf}!n_{BBf}!}\frac{n_m!}{n_{Am}!n_{Bm}!}\frac{\Gamma(\sum_{j=1}^{3} a_{j,3,f})}{\prod_{j=1}^{3}\Gamma(a_{j,3,f})}\frac{\Gamma(\sum_{j=1}^{2} b_{j,3,m})}{\prod_{j=1}^{2}\Gamma(b_{j,3,m})}\times$$

$$\frac{\Gamma(a_{1,3,f}+n_{AAf})\Gamma(a_{2,3,f}+n_{ABf})\Gamma(a_{3,3,f}+n_{BBf})}{\Gamma(\sum_{j=1}^{3} a_{j,3,f}+n_f)}\frac{\Gamma(b_{1,3,m}+n_{Am})\Gamma(b_{2,3,m}+n_{Bm})}{\Gamma(\sum_{j=1}^{2} b_{j,3,m}+n_m)}. \quad (6.4)$$

Note that the only model that requires integration is Model 1. However, it can be carried out numerically without any problem because the integration region is compact, and grid size can be set to be as small as needed for the precision required.

Appendix 2; Posterior distribution under Model 3

Under the saturated Model 3, $(n_{AAf}, n_{ABf}, n_{BBf})$ is Multinomial$(n_f, (p_{AAf}, p_{ABf}, p_{BBf}))$ distributed and n_{Am} is Binomially(n_m, p_{Am}) distributed. Under the assumption that a

priori $(p_{AAf}, p_{ABf}, p_{BBf})$ is Dirichlet$(a_{1,3,f}, a_{2,3,f}, a_{3,3,f})$, and p_{Am} is Beta$(b_{1,3,m}, b_{2,3,m})$, the posterior distribution for $(p_{AAf}, p_{ABf}, p_{BBf})$ is:

$$\pi(p_{AAf}, p_{ABf}, p_{BBf}|y) = \text{Dirichlet}(a_{1,3,f} + n_{AAf}, a_{2,3,f} + n_{ABf}, a_{3,3,f} + n_{BBf}), \tag{6.5}$$

independent of the posterior distribution for p_{Am}, which is:

$$\pi(p_{Am}|y) = \text{Beta}(b_{1,3,m} + n_{Am}, b_{2,3,m} + n_{Bm}). \tag{6.6}$$

The marginal posterior distributions for p_{Af}, f and d follow from the ones for $(p_{AAf}, p_{ABf}, p_{BBf})$ and for p_{Am}, and they can be easily estimated by simulating large samples of $(p_{AAf}, p_{ABf}, p_{BBf})$, and of (p_{Am}, p_{Bm}), and for each value in the sample compute the corresponding value of p_{Af}, of f, and of d, using (2.6), (2.4) and (2.5), respectively.

Acknowledgments

This work was partially supported by grant 2014SGR551 from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) of the Generalitat de Catalunya, by grants MTM2015-65016-C2-2-R (MINECO/FEDER) and MTM2013-43992-R of the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund, and by Grant R01 GM075091 from the United States National Institutes of Health. The authors are extremely grateful for the comments and suggestions for improvement of the associate editor and two referees.

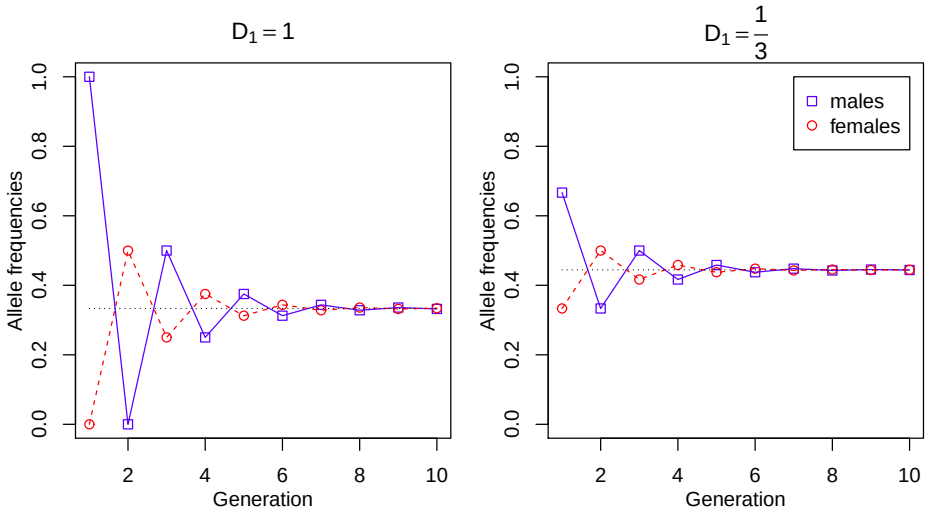

Figure 1: Evolution of allele frequencies over time as a function of the initial difference (D) in allele frequency between males and females. The dotted horizontal line represents the overall A allele frequency. Initial male (p_{Am}) and female (p_{Af}) allele frequencies are (1,0) and (2/3,1/3).

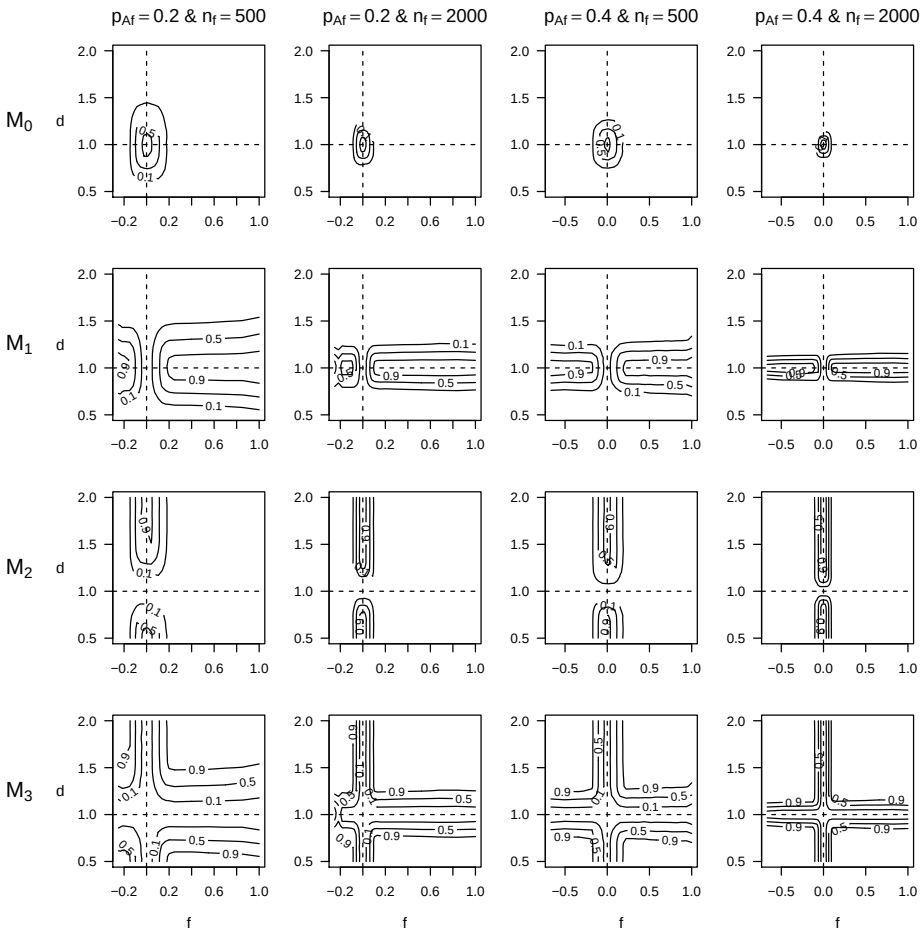

Figure 2: Contour plots for the expected value of $P(M_i|y)$ for $i = 0, 1, 2, 3$ as a function of (f, d), for $p_{Af} = 0.2$ and 0.4 and for $n_f = n_m = 500$ and 2000. The contour levels in all panels are set at .1, .5 and .9.

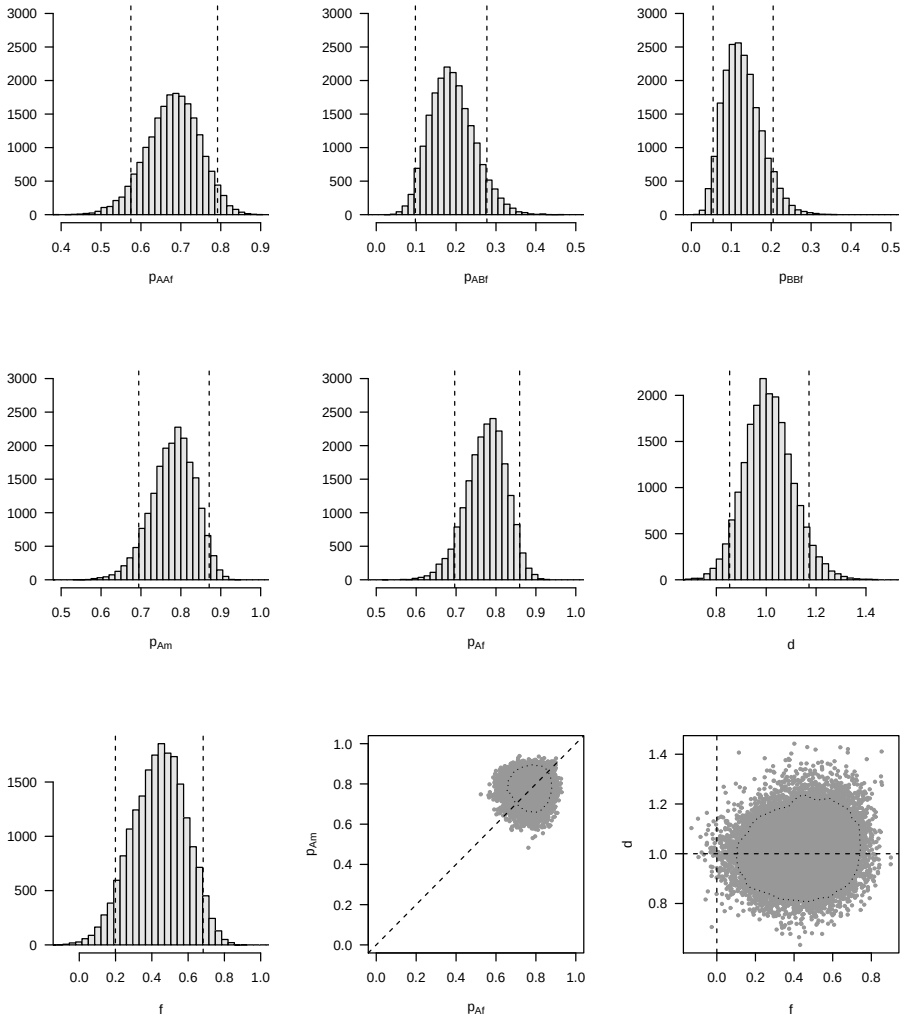

Figure 3: Marginal posterior distributions and 90% hpd posterior credible region of female genotype frequencies, of male and female allele frequencies, of the ratio of male to female allele frequencies and of the inbreeding coefficient, and marginal joint posterior distributions of (p_{Af}, p_{Am}) and of (f, d), all for SNP rs2301322 in Table 1, under the saturated Model.

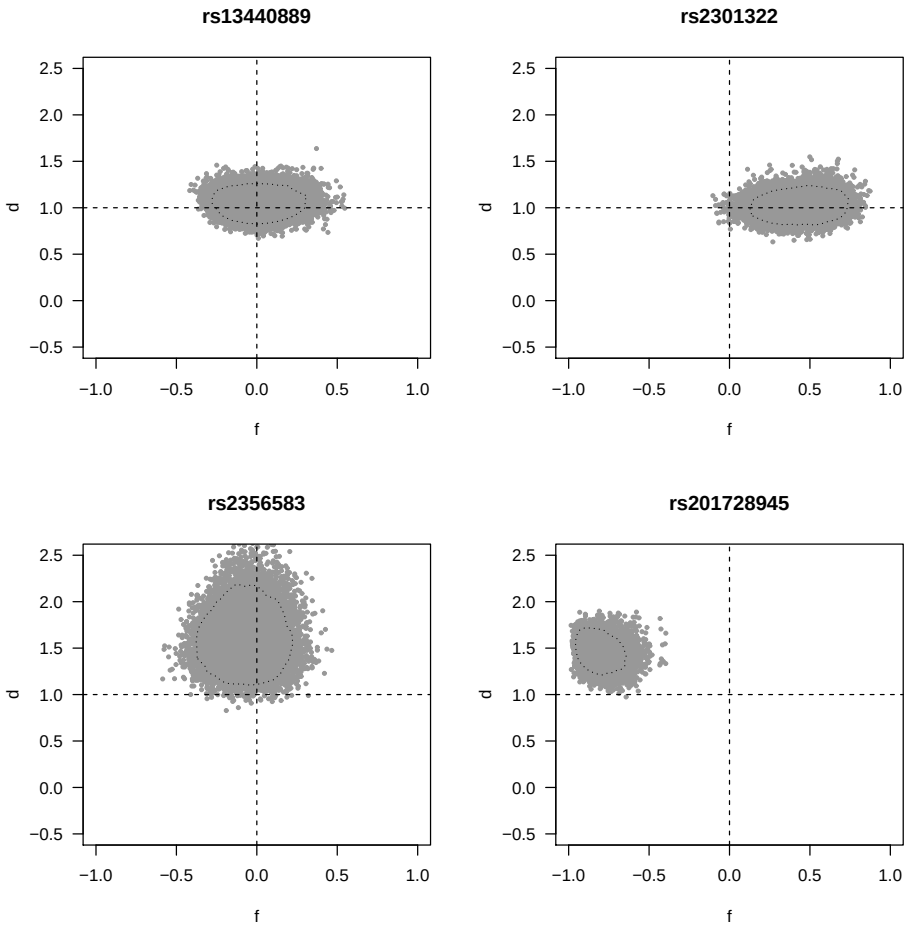

Figure 4: Marginal joint posterior distributions and 90% hpd posterior credible region for (f, d) for the four SNPs considered in Table 1, under the saturated Model.

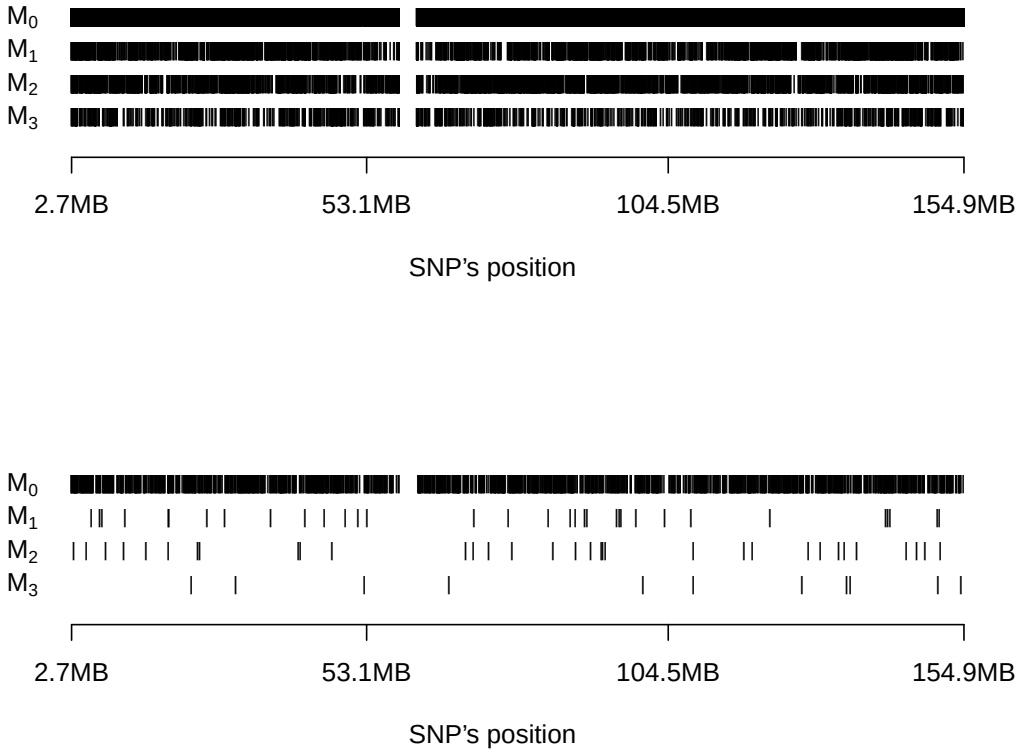

Figure 5: Model with the largest posterior probability, and hence the largest Bayes factor, for the SNPs selected from the Japanese population, presented in the position where they are placed on the X-chromosome. The first panel corresponds to the sample of 162225 SNPs used in Subsection 4.2, and the second one to the 1622 SNPs used in Subsection 4.3. The lower PAR zone is between 60001 and 2699520, and the upper PAR zone between 154931044 and 155260560; no SNPs in these zones were included in the analysis.