Genet. Sel. Evol. 34 (2002) 537–555

DOI: 10.1051/gse:2002022

Original article

Irreducibility and efficiency of ESIP to sample marker genotypes in large pedigrees with loops

Soledad A. Fernández[a], Rohan L. Fernando[b*], Bernt Guldbrandtsen[d], Christian Stricker[e], Matthias Schelling[e], Alicia L. Carriquiry[c]

[a] Department of Statistics, 317 Cockins Hall, Ohio State University, Columbus, OH 43210, USA
[b] Department of Animal Science, 225 Kildee Hall, Iowa State University, Ames, IA 50011, USA
[c] Department of Statistics, 219 Snedecor Hall, Iowa State University, Ames, IA 50011, USA
[d] Danish Institute of Animal Science, Foulum, Denmark
[e] Institute of Animal Sciences, Swiss Federal Institute of Technology, ETH-Zentrum CLU, 8092 Zürich, Switzerland

(Received 21 August 2001; accepted 6 May 2002)

Abstract – Markov chain Monte Carlo (MCMC) methods have been proposed to overcome computational problems in linkage and segregation analyses. This approach involves sampling genotypes at the marker and trait loci. Among MCMC methods, scalar-Gibbs is the easiest to implement, and it is used in genetics. However, the Markov chain that corresponds to scalar-Gibbs may not be irreducible when the marker locus has more than two alleles, and even when the chain is irreducible, mixing has been observed to be slow. Joint sampling of genotypes has been proposed as a strategy to overcome these problems. An algorithm that combines the Elston-Stewart algorithm and iterative peeling (ESIP sampler) to sample genotypes jointly from the entire pedigree is used in this study. Here, it is shown that the ESIP sampler yields an irreducible Markov chain, regardless of the number of alleles at a locus. Further, results obtained by ESIP sampler are compared with other methods in the literature. Of the methods that are guaranteed to be irreducible, ESIP was the most efficient.

Metropolis-Hastings / irreducibility / Elston-Stewart algorithm / iterative peeling

* Correspondence and reprints
E-mail: rohan@iastate.edu

1. INTRODUCTION

QTL mapping includes the estimation of the locations of QTL, of the magnitudes of the QTL effects, and of the frequencies of QTL alleles. When QTL genotypes cannot be observed, marker genotypes are used together with trait phenotypes to map QTL by marker-QTL linkage analysis.

Typically, the mixed model of inheritance is used in linkage analyses. Under this model, the trait is assumed to be influenced by a single QTL linked to a marker (MQTL) and the remaining QTL are assumed to be unlinked to the marker (RQTL). Further, methods and programs (*e.g.* Loki) are also available for multiple QTL. The additive effects of the RQTL are usually assumed to be normally distributed. Under this model the marker-MQTL parameters can be estimated by likelihood or Bayesian approaches.

Both these approaches require computing the likelihood for the model given the observed pedigree, marker genotypes and trait phenotypes. Except for small pedigrees (less than 20 individuals), it is not feasible to compute the exact likelihood for the mixed model of inheritance [1,7,10,11]. Therefore, alternative models have been adopted for which the likelihood can be computed efficiently [1,7,28], or approximations of the likelihood for the mixed model of inheritance are used [10,11,20]. However, these approaches are limited because they cannot easily accommodate more general models.

Markov chain Monte Carlo (MCMC) methods have been proposed to overcome these limitations. In the application of MCMC to likelihood and Bayesian methods, samples are obtained from the conditional distributions, given the observed data, for the missing marker genotypes, the MQTL genotypes, and the additive effects of the RQTL [9,15,31,33]. Further, in the Bayesian approach samples are also obtained from the posterior distribution of the parameters in the model [15,31,33].

The scalar Gibbs sampler provides the easiest method to sample genotypes, where each genotype of an individual is sampled conditional on the genotypes of all the remaining pedigree members. Due to the Markov property of pedigrees [24], the genotype of an individual depends on only its phenotype and the genotypes of its neighbors — parents, spouses, and offspring. Because of this Markov property, the Gibbs sampler is easy to implement. However, Thomas and Cortessis [31] used a hypothetical example to show that when a marker locus has more than two alleles, sampling using the scalar Gibbs sampler may not yield samples from the conditional distribution because the resulting chain may not be irreducible. A chain is said to be irreducible if the probability is nonzero for moving between any two points in the state space in a finite number of steps.

Even when the chain is irreducible, samples may be highly correlated, which is called slow mixing. This is due to the dependence between genotypes

of parents and progeny, with larger progeny groups causing greater dependence [15]. One strategy that was proposed to overcome this problem is the use of blocking Gibbs, which consists of sampling a block of genotypes jointly [15, 17]. Although blocking Gibbs improves mixing, it does not result in a chain that is guaranteed to be irreducible [16]. Ideas to jointly sample the genotypes in complex pedigrees were independently proposed by Heath [13] and Fernández *et al.* [5]. These approaches propose to use an approximate method to obtain candidates that are accepted or rejected by a Metropolis-Hastings step. Heath [13] stated that the approximate peeling method of Thomas [30] seems to be a promising proposal distribution to obtain those candidates. Fernández *et al.* [5] proposed to use a "modified" pedigree as a proposal distribution. This "modified" pedigree is obtained by cutting the loops [29] and extending the pedigree at the cuts [34]. It has been shown that results obtained by "cutting" and "extending" the pedigree can also be obtained by iterative peeling without explicitly modifying the pedigree [34].

Fernández *et al.* [6] implemented a sampling method that combines Elston-Stewart algorithm and iterative peeling, which is called ESIP, to sample genotypes jointly from the entire pedigree. In Fernández *et al.* [6], the mixing properties of ESIP for a trait genotype were examined and documented. In this paper, we show that ESIP results in an irreducible and aperiodic chain even when sampling genotypes at a marker locus with more than two alleles. Here we present a brief description of the method of sampling, a proof that the resulting chain is irreducible and aperiodic, a strategy to improve the efficiency of the sampler, and a comparison of the proposed method with other methods.

2. METHOD FOR SAMPLING GENOTYPES JOINTLY

The method to sample genotypes jointly has been described in detail by Fernández *et al.* [6]. Here, only a brief description is provided to introduce the concepts necessary to prove irreducibility and aperiodicity.

When the pedigree does not have loops or the pedigree contains only simple loops, the entire pedigree is peeled using the Elston-Stewart algorithm [3]. Then genotypes are sequentially sampled using reverse peeling [14, 15, 17]. If the pedigree has complex loops, exact peeling is not feasible [16] and a joint sample is obtained from a pedigree modified to make peeling feasible [6]. This modified pedigree is used to generate the candidates in a Metropolis-Hastings algorithm [12, 23].

This approach to jointly sample marker genotypes is now illustrated with the small pedigree shown in Figure 1a, where the marker genotypes m_3 and m_4 for individuals 3 and 4 are missing.

This pedigree is simple enough to be peeled exactly. However, to illustrate the proposed method the pedigree can be modified as shown in Figure 1b,

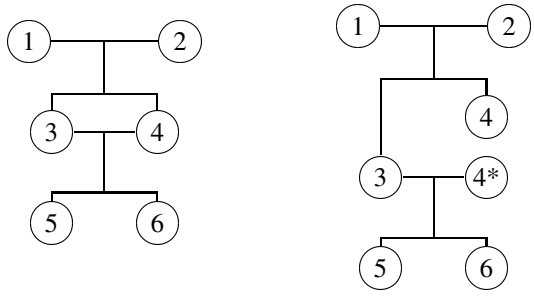

Figure 1. True and cut pedigree, where individuals 1, 2, 5 and 6 have observed marker genotypes.

where individual 4^* has been introduced to remove the loop. This individual is assigned the same genotype as 4, *i.e.*, 4^* is assigned a missing genotype. A pedigree that is modified by duplicating a single individual as shown in Figure 1b will be referred to as a "cut" pedigree. In a cut pedigree, there are two kinds of individuals: those that correspond to individuals in the original pedigree and those that are introduced. Now, the missing genotypes for the original individuals in the cut pedigree can be sampled by reverse peeling [6, 14, 17]. For example in Figure 1b, m_4 is sampled from

$$\Pr(m_4|m_1, m_2, m_5, m_6) = \sum_{m_3} \sum_{m_{4^*}} \Pr(m_3, m_4, m_{4^*}|m_1, m_2, m_5, m_6)$$

which is computed using the Elston-Stewart algorithm [3,6]. Then, m_3 is sampled from

$$\Pr(m_3|m_1, m_2, m_4, m_5, m_6) = \sum_{m_{4^*}} \Pr(m_3, m_4, m_{4^*}|m_1, m_2, m_5, m_6).$$

This gives a joint sample for m_3 and m_4 from $\Pr(m_3, m_4|m_1, m_2, m_5, m_6)$. In general, the missing genotypes for the original individuals are sampled conditional on the observed genotypes. This sample from the cut pedigree is either accepted or rejected according to Metropolis-Hastings algorithm as described below.

We use a special case of the Metropolis-Hastings algorithm known as the independence sampler. Let $\mathbf{y}$ be the vector of observed genotypes and $\mathbf{m}$ the vector of missing genotypes. In this algorithm, the candidate draw is accepted with probability

$$\eta(\mathbf{m}_{prev}, \mathbf{m}_c) = min\left(1, \frac{\pi(\mathbf{m}_c)q(\mathbf{m}_{prev})}{\pi(\mathbf{m}_{prev})q(\mathbf{m}_c)}\right), \tag{1}$$

where $\pi(\mathbf{x})$ is the probability of sampling $\mathbf{x}$ from the pedigree in Figure 1a conditional on $\mathbf{y}$, $q(\mathbf{x})$ is the probability of sampling $\mathbf{x}$ from the pedigree in Figure 1b conditional on $\mathbf{y}$, $\mathbf{m}_c$ is the candidate sample obtained from the pedigree in Figure 1b, and $\mathbf{m}_{prev}$ is the last vector of genotypes that was accepted.

In general, the probability $\pi(\mathbf{m})$ can be computed as

$$\pi(\mathbf{m}) \propto \prod_{j=1}^{n} \pi_j, \tag{2}$$

where

$$\pi_j = \begin{cases} \Pr(m_j) & \text{if } j \text{ is a founder} \\ \Pr(m_j | m_{m_j}, m_{f_j}) & \text{if } j \text{ is an offspring.} \end{cases}$$

To compute $q(\mathbf{m})$ we multiply the probabilities that were used in the sampling process. For this example, $q(\mathbf{m})$ is

$$q(\mathbf{m}) = \Pr(m_4|\mathbf{y}) \Pr(m_3|m_4, \mathbf{y}). \tag{3}$$

2.1. Proof of irreducibility and aperiodicity

Let I be the state space for the vector of unobserved genotypes in the unmodified pedigree, and let $\mathbf{m}_i$ and $\mathbf{m}_j$ be two arbitrary states from I. The Markov chain for sampling genotypes is irreducible if the probability of moving from $\mathbf{m}_i$ to $\mathbf{m}_j$ in a finite number of steps is nonzero. We show below that for the ESIP sampler, the probability of going from $\mathbf{m}_i$ to $\mathbf{m}_j$ in one step is nonzero. This probability of going from $\mathbf{m}_i$ to $\mathbf{m}_j$ is

$$\begin{aligned} \Pr(\mathbf{m}_j|\mathbf{m}_i) &= \eta(\mathbf{m}_i, \mathbf{m}_j)\, q(\mathbf{m}_j) \\ &= min\left(1, \frac{\pi(\mathbf{m}_j) q(\mathbf{m}_i)}{\pi(\mathbf{m}_i) q(\mathbf{m}_j)}\right) q(\mathbf{m}_j). \end{aligned} \tag{4}$$

Note that $\pi(\mathbf{m}_i) > 0$ and $\pi(\mathbf{m}_j) > 0$ because $\mathbf{m}_i$ and $\mathbf{m}_j$ are in I. Further, as shown in the Appendix, if $\pi(\mathbf{m}) > 0$ then $q(\mathbf{m}) > 0$. So in (4), $\eta(\mathbf{m}_i, \mathbf{m}_j) > 0$ and $q(\mathbf{m}_j) > 0$, and thus $\Pr(\mathbf{m}_j|\mathbf{m}_i) > 0$. This shows that the chain has a nonzero probability of moving from any state $\mathbf{m}_i$ to any other state $\mathbf{m}_j$ in a single step. Thus, this proves that the chain is irreducible and aperiodic.

3. IMPROVING EFFICIENCY

Sampling genotypes as described above can be inefficient in a pedigree with many loops. To illustrate, consider the case of a biallelic marker locus with alleles M_1 and M_2. In the pedigree in Figure 1a, the marker genotypes

of individuals 3 and 4 are unobserved. To sample genotypes we introduce individual 4* to remove the loop (Fig. 1b). Assume that the genotypes of 1, 2, 5 and 6 are M_1M_2, M_1M_2, M_1M_1 and M_1M_2 respectively. Now, to sample m_3 we use $\Pr(m_3|\mathbf{y})$. Next, we sample m_4 using $\Pr(m_4|\mathbf{y}, m_3) = \Pr(m_4|m_1, m_2)$. Now that both unknown genotypes have been sampled, we computed $q(\mathbf{m}_c)$ as

$$q(\mathbf{m}_c) = \Pr(m_3|\mathbf{y}) \Pr(m_4|m_1, m_2).$$

To compute η we also need $q(\mathbf{m}_{prev})$. This quantity has already been calculated from a previous round of the sampler. Further, we need the probabilities $\pi(\mathbf{m}_c)$ of the candidate sample $\mathbf{m}_c$ and $\pi(\mathbf{m}_{prev})$ of the accepted sample $\mathbf{m}_{prev}$ from the previous round. Computing $\pi(\mathbf{m}_c)$ is straightforward using (2). Again, $\pi(\mathbf{m}_{prev})$ has already been computed in the previous round of sampling.

Suppose that m_3 was sampled as M_2M_2 and m_4 as M_2M_2. Then $\mathbf{m}_c{}' = (M_2M_2, M_2M_2)$ and $\pi(\mathbf{m}_c) = 0$ because individual 4 with genotype M_2M_2 cannot have offspring 5 with genotype M_1M_1. As a result $\eta = 0$ and the candidate sample will be rejected with probability 1. We showed earlier that $\pi(\mathbf{m}_c) > 0$ implies $q(\mathbf{m}_c) > 0$, but this example shows that $q(\mathbf{m}_c) > 0$ does not imply $\pi(\mathbf{m}_c) > 0$. The probability of getting a candidate rejected increases with the number of loops.

One strategy to improve efficiency of the sampler is to minimize the number of loops that are cut. When peeling is applied to a pedigree, intermediate results are stored in multidimensional tables called "cutsets" [2]. In a pedigree without loops, the largest cutset has dimension two. In a pedigree with loops, some cutsets have dimension greater than two. Depending on the pedigree, peeling can be efficient as long as the dimension of the largest cutset is about seven [6]. In the ESIP sampler, exact peeling is applied until the cutset size is too large for efficient computations. To proceed further, loops are cut.

A second strategy to improve efficiency of the sampler consists of extending the pedigree at the places it was cut. Wang *et al.* [34] have shown that the approximation to the likelihood obtained by cutting loops is improved when the pedigree is extended as shown in Figure 2. So, it seems reasonable to expect that cutting loops and extending the pedigrees will also reduce the probability of getting a candidate rejected. In Figure 2 the pedigree is extended by including individuals 5* and 6* as offspring of individuals 4 and 3*. A pedigree modified by duplicating more than a single individual will be referred to as a "cut-extended" pedigree.

The probabilities of getting a rejected sample were obtained for the cut pedigree shown in Figure 1b and for the cut-extended pedigree shown in Figure 2. As before, it was assumed that individuals 1,2,5 and 6 have genotypes M_1M_2, M_1M_2, M_1M_1 and M_1M_2, respectively. The gene frequencies were assumed to be 0.5 for each allele. The probabilities of getting a rejected sample were 0.333 for the cut pedigree and 0.111 for the cut-extended pedigree.

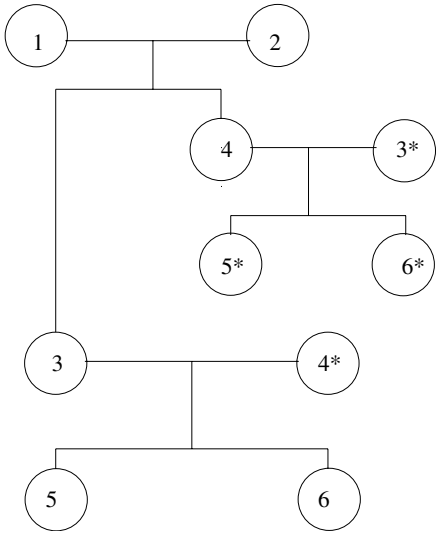

Figure 2. Cut-extended pedigree. Marker genotypes were observed for individuals 1, 2, 5 and 6. If the genotype of individual i is observed, the extended individual i^* is assigned the same genotype as individual i.

"Cutting" and "extending" the pedigrees is difficult and the degree of difficulty increases as the loops are larger and more complex. In practice, the pedigree does not have to be extended explicitly. In Wang *et al.* [34] it was shown that genotype probabilities computed by iterative peeling are equivalent to genotype probabilities computed from a cut-extended pedigree. As explained in Fernández *et al.* [6], the ESIP sampler combines the Elston-Stewart algorithm and iterative peeling to sample genotypes jointly from the entire pedigree.

To speed up peeling, genotype elimination was implemented using the algorithm developed by Lange and Goradia [19]. This algorithm is an extension of Lange and Boehnke [18] and consists of identifying all those genotypes that are not consistent with the observed information in the pedigree. These genotypes have zero probability and are removed from the list of genotypes to be summed over in peeling.

4. PERFORMANCE OF THE ESIP SAMPLER

To assess the performance of ESIP we have compared its efficiency and accuracy with those of other MCMC methods proposed in the literature. One of the methods that is guaranteed to yield an irreducible chain is given by Sheehan and Thomas [24]. In this paper this method will be referred to as the Sheehan-Thomas sampler. Lin *et al.* [22] and Lin [21] have also proposed two methods for sampling marker genotypes. These will be referred to as

Lin1 and Lin2 samplers, respectively. Sobel and Lange [25] have described how samples of descent graphs can be used for linkage analysis rather than samples of descent states. It has been argued that the space of descent graphs is much smaller than the space of descent states. However for comparison with ESIP, as described in Section 5, genotype probabilities can be estimated from a sample of descent graphs. This method will be referred to as the Descent-graph sampler.

4.1. Comparison of ESIP and Sheehan-Thomas samplers

Regardless of the number of the alleles at a locus, Sheehan and Thomas [24] have shown that if all penetrance probabilities are non-zero then irreducibility holds. Let $\pi^*(\mathbf{m})$ be the distribution of $\mathbf{m}$ given $\mathbf{y}$ after all zero penetrance probabilities have been replaced by some small positive probability (relaxation parameter). They showed that if samples are obtained from $\pi^*(\mathbf{m})$ and those for which $\pi(\mathbf{m}) = 0$ are rejected, then the remaining samples are from $\pi(\mathbf{m})$. Thus, to overcome the irreducibility problem they proposed to sample from $\pi^*(\mathbf{m})$ and only use samples for which $\pi(\mathbf{m}) > 0$ to estimate genotype probabilities.

They also showed that if all transmission probabilities are non-zero irreducibility holds. So, an alternative $\pi^*(\mathbf{m})$ to sample missing genotypes from can be obtained by modifying the transmission probabilities and/or penetrance probabilities.

Sheehan and Thomas [24] estimated genotype probabilities by their method for the *ABO* blood type locus in the fictitious pedigree given in [24] (Fig. 1).

In this pedigree, squares represent males and circles represent females. The *ABO* blood-group system consists of three alleles *A*, *B* and *O*, and hence six genotypes. However, there are only four phenotypes, as only *A* and *B* are codominant, and both, *A* and *B*, are dominant to *O*. Thus, the *AA* and *AO* genotypes are phenotypically indistinguishable and give blood type *A*; similarly, the *BB* and *BO* give blood type *B*. The *O* blood group corresponds only to the recessive genotype *OO*; while *AB* genotypes are distinguishable from other genotypes. Six individuals in the pedigree shown in [24] (Fig. 1) have genetic data (12 and 21 have genotypes *AB*; 16, 17, 18 and 19 have genotype *OO*). As Sheehan and Thomas [24] explained, these phenotypes were deliberately chosen so that the mated pair [6,9] could be either (*AO*, *BO*) or (*BO*, *AO*) and these two states do not communicate. The same applies to the pair [10,15]. The assumed allele frequencies for *A*, *B* and *O* alleles are 0.2, 0.1 and 0.7, respectively. Even though this pedigree has loops, it is small enough that exact marginal probabilities can be calculated for all individuals.

Results obtained by the ESIP and Sheehan-Thomas samplers were compared to the true marginal probabilities. Sheehan and Thomas [24] explained that there is a trade-off between the size of the relaxation parameter and efficiency

of the algorithm. If the relaxation parameter is too small then the Markov chain has slow mixing because stepping between non-communicating classes has too small a probability. On the other hand, if the relaxation parameter is too large too many samples will be rejected. They presented results for some individuals in the pedigree using different relaxation parameters. Based on those results the value of 0.025 was chosen for the relaxation parameter to estimate genotype probabilities for the entire pedigree.

Different versions of the ESIP sampler were used to compare with results obtained by Sheehan and Thomas [24]. The first version, which is called *Direct*, consists of peeling exactly the whole pedigree and then samples are obtained directly from the target distribution by reverse peeling. When the proposal is obtained by exactly peeling the pedigree until the cutset size is k and then iterative peeling is applied to the remainder, the sampler is called ESIP-k. For this pedigree, $k = 2$ and $k = 3$ were also used for comparison. The length of the chain for the three cases (*Direct*, ESIP-3, and ESIP-2) was 10 000 with no burn-in period.

The mean difference between Sheehan-Thomas sampler and the true marginal probabilities is 1.8×10^{-3}, and the largest difference is 1.1×10^{-2}. The total number of simulations for the Sheehan-Thomas sampler was 175 830 with a rejection rate of 94.31%, which yields a total of 10 000 legal samples. Also, genotype probabilities were obtained by the *Direct*, ESIP-2, and ESIP-3 samplers and compared to the true marginal probabilities. Detailed tables that show the difference mean, range and standard deviation by genotype are given in Fernández [4]. The mean difference with the *Direct* sampler is 1.6×10^{-3}, and the largest difference is 1.1×10^{-2}. The mean difference with the ESIP-2 sampler is 1.9×10^{-3} and the largest difference is 1.2×10^{-2}. The rejection rate for this sampler was 24.5%. For ESIP-3 (with 10 000 samples), the mean difference is 1.4×10^{-3} and the largest difference is 1.1×10^{-2}. These values are the same as the results obtained for the *Direct* sampler. The rejection rate for ESIP-3 was 6.5%. These differences show that the ESIP sampler yields results with the same level of accuracy than Sheehan-Thomas sampler. Also, the rejection rates for the ESIP sampler are much lower than Sheehan-Thomas sampler.

The accuracy of the estimates obtained by ESIP greatly improve as the number of samples is increased. For ESIP-3, the mean differences are 3.1×10^{-4} and 1.2×10^{-4}, for chain lengths of 100 000 and 1 000 000, respectively. The largest differences are 3.1×10^{-3} and 1.6×10^{-3}, respectively. The accuracy of the Sheehan-Thomas sampler may not increase when the number of samples is increased because it is well known that Gibbs sampler has slow mixing [6, 8, 15, 17].

ESIP was run using a Pentium Pro-200. The computing times were 90, 36 and 12 s for ESIP-2, ESIP-3, and *Direct*, respectively. Sheehan and

Thomas [24] used a SUN SPARC station SLC and the reported computing time is 344.64 s. But, it is difficult to compare the computing times of ESIP and Sheehan-Thomas because different computing systems were used. However, as explained below, for a single locus, the number of samples can be used for comparison instead of using the computing times.

The computing time for ESIP can be split into two components: peeling time and sampling time. Relative to sampling time, peeling time is negligible because it is done only once. Further, for an exactly peeled individual, the computations needed to sample the genotype by reverse peeling are very similar to the computations in the Gibbs sampler [6]. Thus, the number of samples from the *Direct* sampler are directly comparable to the number of samples from Sheehan-Thomas sampler, which is based on the Gibbs sampler. For this pedigree, the *Direct* sampler with a chain length of 10 000 yields the same level of accuracy than the Sheehan-Thomas sampler with 175 830 simulations. Therefore, the *Direct* sampler is more efficient.

As explained below, the number of samples from ESIP when iterative peeling is applied to a part of the pedigree, cannot be directly compared with the number of samples from the Sheehan-Thomas sampler. For the ESIP-k sampler, when an individual that was iteratively peeled has to be sampled, all the cutsets connected to this individual must be recalculated conditional on the genotypes that have already been sampled [6]. This can be very time consuming because iteratively peeled individuals are connected to cutsets that contain a mixture of individuals that are sampled and not sampled. Thus, this recalculation involves summing over all genotypes of the individuals that were not yet sampled conditional on the genotypes that have been already sampled. This process has to be repeated in each sample. On the contrary, when individuals are peeled exactly, all the other individuals in cutsets connected to the individual being sampled have already been sampled. Thus, there is no summing over that needs to be done. This indicates that a large improvement in the efficiency of the ESIP sampler will be possible if all loops in the pedigree are cut when the cutset size of k is reached. After cutting, exact peeling can be applied to obtain samples more efficiently. Briefly, exact peeling is first applied until cutset size is k. Second, iterative peeling is applied to the remaining individuals in the pedigree. Third, all loops in the pedigree are cut. Fourth, exact peeling is continued using the iteratively peeled probabilities where the loops were cut. As shown by Wang *et al.* [34] this is equivalent to cutting and extending the pedigrees at the cuts.

4.2. Comparison of ESIP and Lin1 samplers

Lin *et al.* [22] presented results obtained by the application of their method in a Volga German family to study Alzheimer's disease. The marker locus for the Alzheimer's disease (D14S43) has three alleles: A, B and C. The frequencies

they used were 0.239, 0.760 and 0.001 for the three alleles, respectively. Lin *et al.* [22] presented results for nine individuals of the pedigree (shown in Fig. 3, [22]).

In the Lin1 sampler, marker genotypes are sampled using the scalar-Gibbs sampler. As described below, the irreducibility problem is overcome by coupling an auxiliary Markov chain that is irreducible with the scalar-Gibbs chain [22].

Let Γ_θ be the scalar-Gibbs chain with equilibrium distribution $P_\theta(g_\theta|\mathbf{y})$, where g_θ denotes a genotypic configuration in the state space G_θ of the scalar-Gibbs sampler. Similarly, $\Gamma_{\theta'}$ is the irreducible-auxiliary chain with equilibrium distribution $P_{\theta'}(g_{\theta'}|\mathbf{y})$, where $g_{\theta'}$ denotes a genotypic configuration in the state space $G_{\theta'}$ of the auxiliary chain. These two chains are coupled by switching their states. If an appropriate switching probability is used, it has been shown that the coupled chain Γ^* defined on the state space $G_\theta \times G_{\theta'}$ is irreducible and has equilibrium distribution $P_\theta(g_\theta|\mathbf{y})P_{\theta'}(g_{\theta'}|\mathbf{y})$. Thus, the $\{(g_\theta^i)\}$ component of the coupled chain converges to $P_\theta(g_\theta|\mathbf{y})$ [22].

Lin *et al.* [22] showed that for a scalar-Gibbs sampler the chain is irreducible if and only if, each heterozygote genotype, has a positive penetrance probability. In the Lin1 sampler, the auxiliary chain $\Gamma_{\theta'}$ was constructed by setting each heterozygote genotype A_mA_n to have a small positive penetrance probability ρ_{mn}. If ρ_{mn} is too small, the probability of switching is too small. On the other hand, if ρ_{mn} is too large, many of the $g_{\theta'}$ will have zero probability in the state space G_θ, resulting in the switches being rejected. To overcome this, the heated Metropolis algorithm was used to simulate the auxiliary chain. Because a single heated auxiliary chain did not improve mixing in some cases, Lin *et al.* [22] used multiple auxiliary chains.

The mean difference of the results presented in by Lin *et al.* [22] and the true marginal probabilities is 9.4×10^{-4} and the largest difference is 6.0×10^{-3}. The estimates were obtained from 400 000 samples using three auxiliary chains. Thus, this requires generating four chains, each of length 400 000.

For comparison, a chain length of 20 000 with no burn-in period was used for all the ESIP samplers. Detailed tables that show the mean, range and standard deviation of the difference between the ESIP samplers and the true marginal probabilities, by genotype, are given in Fernández [4]. For the *Direct* sampler, the mean difference is 7.3×10^{-4} and the largest difference is 2.2×10^{-3}. This indicates that this sampler yields results more accurate than the Lin1 sampler. In addition, the number of samples required to obtain this level of accuracy using the ESIP sampler is much smaller than the samples required in the Lin1 sampler.

For the ESIP-2 sampler, the mean difference is 1.1×10^{-3} and the largest difference is 4.9×10^{-3}. The rejection rate for this sampler was 23.86%. These results have the same level of accuracy as the Lin1 sampler.

For the ESIP-3 sampler, the mean difference is 1.0×10^{-3} and the largest difference is 6.3×10^{-3}. The rejection rate for this sampler was 15.25%. Thus, the level of accuracy for this sampler with 20 000 samples is the same as the Lin1 sampler with 400 000 samples using three auxiliary chains.

The Lin1 sampler samples one variable at a time from the full conditional, so the number of samples from Lin1 and *Direct* gives a good measure of efficiency. The same level of accuracy was obtained from the *Direct* sampler with a chain length of 20 000 and the Lin1 sampler with 400 000 samples. Thus, the *Direct* sampler is more efficient.

Furthermore, as Lin *et al.* [22] explained, their approach may not be practical when a locus has more than three alleles because there are a larger number of non-communicating classes. This is not a problem for ESIP.

4.3. Comparison of ESIP and Lin2 samplers

The estimates obtained by the ESIP sampler were also compared to those obtained by Lin [21]. She used the same *ABO*-blood-type pedigree used by Sheehan and Thomas [24] to show the performance of her method. She proposed a method where an irreducible chain is constructed by jumping from one communicating class to another directly without the need of stepping through illegal configurations. This method also requires the explicit identification of non-communicating classes.

Lin [21] estimated genotype probabilities from a chain of length 3 000 cycles. For comparison, a chain length of 3 000 with no burn-in period was used for different ESIP samplers (*Direct*, ESIP-2 and ESIP-3). The same level of accuracy as for the Lin2 sampler was obtained with the ESIP samplers.

As Lin [21] explained, her algorithm is efficient as long as one can identify individuals in the pedigree who characterize the non-communicating classes. Her method is not a single component algorithm, since the first step is to identify all the non-communicating classes. Thus, as Lin [21] added, one can design blocking Gibbs sampling algorithms to accomplish the same task. The reason is that the identification of all non-communicating classes is also the basis of designing blocking Gibbs samplers.

On the contrary, for the chain generated by the ESIP sampler, all states communicate, and thus, irreducibility is guaranteed. The performance of ESIP sampler was also tested using large and complex pedigrees. For more details see Fernández *et al.* [6].

5. COMPARISON OF ESIP AND DESCENT-GRAPH SAMPLERS

The estimates obtained by the ESIP sampler were also compared to those obtained by the Descent graph sampler developed by Sobel and Lange [25].

To estimate probabilities on pedigrees, Thompson [32] proposed an alternative MCMC strategy, where segregation indicators are sampled rather than genotypes. She argued that the advantage of this method is that the space of segregation indicators is much smaller than the space of genotypes, especially for multiallelic loci. Sobel and Lange [25] have implemented such a sampler. This sampler will be referred to as the Descent-graph sampler. Results from the Descent-graph sampler can be used to estimate genotype probabilities. Here, we used the Descent-graph sampler to obtain genotype probabilities for the *ABO*-blood-type pedigree used by Sheehan and Thomas [24]. The results obtained from this sampler are compared to the true marginal probabilities. The Descent-graph and *Direct* samplers were run on the same system to compare the computing times. The computing times for 100 00 samples were 2 400 s and 250 s for Descent-graph and *Direct*, respectively. Thus, *Direct* is about 10 times faster than Descent-graph. We also ran ESIP-2 and ESIP-3 for 100 000 samples. The computing times were 1 320 and 660 seconds for ESIP-2 and ESIP-3, respectively.

A chain of 10 000 samples was used to obtain estimates from the Descent-graph sampler. The absolute mean difference across genotypes between the estimates and the true marginal probabilities is 3.6×10^{-3}. The same level of accuracy was obtained form the *Direct* sampler with a chain of only 500 samples. The mean absolute difference for *Direct* was 6.2×10^{-3}. The largest absolute difference for the Descent-graph sampler is 6.2×10^{-2}, and the largest absolute difference for the *Direct* sampler is 5.8×10^{-2}. Thus, these results show that the *Direct* sampler yields estimates with the same level of accuracy as the Descent-graph sampler in much less time. ESIP-2 and ESIP-3 yielded similar results to those obtained by *Direct*.

The Descent-graph sampler was also used to estimate probabilities for a biallelic locus in a large half-sib family. The allele frequencies are 0.75 and 0.25 for allele *A* and *a*, respectively. The pedigree consists of three founders: one sire and two dams, where each family has 35 offspring. In both nuclear families, the genotype for 34 of the offspring is known, 17 are homozygous *AA* and 17 heterozygous *Aa*. The genotype of the parents and one offspring in each nuclear family is unknown. Four different initial descent graphs (Descent-graph$^{(1)}$, Descent-graph$^{(2)}$, Descent-graph$^{(3)}$ and Descent-graph$^{(4)}$) were used to obtain estimates for this pedigree. In all the initial Descent graphs, the six founder alleles are labeled as: paternal alleles (3,4) and maternal alleles (1,2 and 5,6). The two families are labeled 1 and 2. Family 1 has founder alleles 1,2 and 3,4; and family 2 has founder alleles 5,6 and 3,4. In Descent-graph$^{(1)}$: the *AA* offspring of family 1 inherited founder alleles 2 and 4, and the *Aa* offspring of family 1 inherited founder alleles 1 and 4; the *AA* offspring of family 2 inherited founder alleles 6 and 4, and the *Aa* offspring of family 2 inherited founder alleles 5 and 4. In Descent-graph$^{(2)}$: the *AA* offspring of family 1

Table I. Estimated marginal probabilities obtained by the Descent-graph and ESIP samplers, and exact marginal probabilities obtained by SALP for two individuals with unknown genotype of a large half-sib family.

Individual	Method	$P(AA)$	$P(Aa)$	$P(aa)$
Parent	SALP	0.599999	0.400001	0.0
	ESIP	0.595	0.405	0.0
	Descent-graph$^{(1)}$	0.573753	0.426247	0.0
	Descent-graph$^{(2)}$	0.57298	0.42702	0.0
	Descent-graph$^{(3)}$	0.881964	0.118036	0.0
	Descent-graph$^{(4)}$	0.0	1.0	0.0
Offspring	SALP	0.499999	0.500000	0.000001
	ESIP	0.4998	0.5002	0.0
	Descent-graph$^{(1)}$	0.461184	0.538816	0.0
	Descent-graph$^{(2)}$	0.540265	0.459735	0.0
	Descent-graph$^{(3)}$	0.512912	0.487088	0.0
	Descent-graph$^{(4)}$	0.2425	0.5071	0.2504

Estimates by ESIP are from 10 000 samples.
Estimates by Descent-graph are from 1 000 000 samples.

inherited founder alleles 1 and 4, and the *Aa* offspring of family 1 inherited founder alleles 1 and 3; the *AA* offspring of family 2 inherited founder alleles 5 and 4, and the *Aa* offspring of family 2 inherited founder alleles 5 and 3. In Descent-graph$^{(3)}$: the *AA* offspring of family 1 inherited founder alleles 1 and 4, and the *Aa* offspring of family 1 inherited founder alleles 1 and 3; the *AA* offspring of family 2 inherited founder alleles 6 and 4, and the *Aa* offspring of family 2 inherited founder alleles 5 and 4. In Descent-graph$^{(4)}$: the *AA* offspring of family 1 inherited founder alleles 2 and 4, and the *Aa* offspring of family 1 inherited founder alleles 2 and 3 or 1 and 4; the *AA* offspring of family 2 inherited founder alleles 6 and 4, and the *Aa* offspring of family 2 inherited founder alleles 6 and 3 or 5 and 4.

Genotype probabilities were also estimated by ESIP, and exactly calculated by SALP [26,27]. Results for two individuals with unknown genotype (one parent and one offspring) are presented in Table I.

This example illustrates that Descent-graph sampler does not have good mixing properties for some pedigrees. For this pedigree, estimates based on 1 000 000 samples from Descent-graph$^{(1)}$ and Descent-graph$^{(2)}$ seem to converge to the true marginal probabilities only for the offspring with unknown genotype. Descent-graph$^{(3)}$ and Descent-graph$^{(4)}$ do not converge to the true marginal probabilities for any of the individuals with unknown genotype. On the other hand, the ESIP sampler converges to the true probabilities with only 10 000 samples.

6. SUMMARY AND CONCLUSIONS

In this paper we have showed that the ESIP sampler is aperiodic and irreducible. We also have compared the ESIP sampler with other samplers in the literature. The ESIP sampler seems to be more efficient than Sheehan-Thomas, Lin1 and Descent-graph samplers. For the same level of accuracy, the ESIP sampler needed much less samples than the the Sheehan-Thomas, Lin1 and Descent-graph samplers. These samplers are guaranteed to give irreducible chains. ESIP seems to be as efficient as the Lin2 sampler. They have the same accuracy in about the same number of samples, but the Lin2 sampler requires identifying non-communicating classes, which may be impossible in large pedigrees.

The Sheehan-Thomas and Lin1 samplers have addressed the irreducibility problem of the scalar-Gibbs sampler, but those samplers still are very inefficient. As Geyer and Thompson [8] indicated, methods that sample one variable at a time, like scalar-Gibbs sampler, can take long time to obtain a representative sample of genotypic configurations. This problem is even more evident in large pedigrees because the time increases exponentially with the number of individuals in the pedigree. This is not a problem for the ESIP sampler, which updates the genotypes jointly. Furthermore, the ESIP sampler has been tested in large pedigrees and it seems to perform well [6].

The Descent-graph sampler of Sobel and Lange [25] was not designed for computing genotype probabilities, but in this paper we used it to obtain genotype probabilities to compare with ESIP. We have shown that ESIP is more efficient and also that the Descent-graph sampler has poor mixing properties for some pedigrees.

In this paper we have examined the properties of ESIP when sampling genotypes at a single locus. ESIP can sample genotypes jointly at multiple linked loci, but this may be inefficient. A better strategy would be to sample genotypes at one locus conditional on other loci, but this method will have horizontal dependence problems. Thus, strategies to overcome horizontal dependence need to be examined.

ACKNOWLEDGEMENTS

The authors are grateful to Professor Wolfgang Kliemann for reviewing the proof of irreducibility and aperiodicity presented in this paper.

REFERENCES

[1] Bonney G.E., Compound regressive models for family data, Hum. Hered. 42 (1992) 28–41.

[2] Cannings C., Thompson E.A., Skolnick M.H., Probability functions on complex pedigrees, Adv. Appl. Prod. 10 (1978) 26–61.

[3] Elston R.C., Stewart J., A general model for the genetic analysis of pedigree data, Hum. Hered. 21 (1971) 523–542.
[4] Fernández S.A., An algorithm to sample genotypes in complex pedigrees, Ph.D. thesis, Iowa State University, 2001.
[5] Fernández S.A., Fernando R.L., Carriquiry A.L., An algorithm to sample marker genotypes in a pedigree with loops, in: Proc. of the American Statistical Association, Section on Bayesian Statistical Science, Alexandria, VA, 1999, pp. 60–65.
[6] Fernández S.A., Fernando R.L., Guldbrandtsen B., Totir L.R., Carriquiry A.L., Sampling genotypes in large pedigrees with loops, Genet. Sel. Evol. 33 (2001) 337–367.
[7] Fernando R.L., Stricker C., Elston R.C., The finite polygenic mixed model: an alternative formulation for the mixed model of inheritance, Theor. Appl. Genet. 88 (1994) 573–580.
[8] Geyer C.J., Thompson E.A., Annealing Markov chain Monte Carlo with applications to ancestral inference, J. Am. Stat. Assoc 90 (1995) 909–920.
[9] Guo S.W., Thompson E.A., A Monte Carlo method for combined segregation and linkage analysis, Am. J. Hum. Genet. 51 (1992) 1111–1126.
[10] Hasstedt S.J., A mixed model approximation for large pedigrees, Comput. Biomed. Res. 15 (1982) 195–307.
[11] Hasstedt S.J., A variance components/major locus likelihood approximation on quantitative data, Genet. Epidemiol. 8 (1991) 113–125.
[12] Hastings W.K., Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109.
[13] Heath S.C., Markov Chain Monte Carlo segregation and linkage analysis for oligogenic models, Am. J. Hum. Genet. 61 (1997) 748–760.
[14] Heath S.C., Generating consistent genotypic configurations for multi-allelic loci and large complex pedigrees, Hum. Hered. 48 (1998) 1–11.
[15] Janss L.L.G., Thompson R., Van Arendonk J.A.M., Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations, Theor. Appl. Genet. 91 (1995) 1137–1147.
[16] Jensen C.S., Kong A., Blocking Gibbs sampling for linkage analysis in large pedigrees with many loops, Am. J. Hum. Genet. 65 (1999) 885–901.
[17] Jensen C.S., Kong A., Kjærulff U., Blocking Gibbs sampling in very large probabilistic expert systems, Int. J. Hum. Comp. Stud. 42 (1995) 647–66.
[18] Lange K., Boehnke M., Extensions to pedigree analysis. V. Optimal calculation of Mendelian likelihoods, Hum. Hered. 33 (1983) 291–301.
[19] Lange K., Goradia T.M., An algorithm for automatic genotype elimination, Am. J. Hum. Genet. 40 (1987) 250–256.
[20] LeRoy P., Elsen J.M., Knott S., Comparison of four statistical methods for detection of a major gene in a progeny test design, Genet. Sel. Evol. 21 (1989) 341–357.
[21] Lin S., A scheme for constructing an irreducible Markov chain for pedigree data, Biometrics 51 (1995) 318–322.
[22] Lin S., Thompson E., Wijsman E., Achieving irreducibility of the Markov chain Monte Carlo method applied to pedigree data, IMA J. Math. Appl. Med. Biol. 10 (1993) 1–17.

[23] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., Equation of state calculation by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092.
[24] Sheehan N., Thomas A., On the irreducibility of a Markov Chain defined on a space of genotype configurations by a sampling scheme, Biometrics 49 (1993) 163–175.
[25] Sobel E., Lange K., Descent graphs in pedigree analysis: Applications to haplotyping location scores, and marker-sharing statistics, Am. J. Hum. Genet. 58 (1996) 1323–1337.
[26] Stricker C., Fernando R.L., Elston R.C., SALP — segregation and linkage analysis for pedigrees, in: Proc. 5th World Cong. Genet. Appl. Livest. Prod., Guelph, August 7–12, 1994a, University of Guelph.
[27] Stricker C., Fernando R.L., Elston R.C., SALP — segregation and linkage analysis for pedigrees, release 2.0, computer program package, 1994b.
[28] Stricker C., Fernando R.L., Elston R.C., Linkage analysis with an alternative formulation for the mixed model of inheritance: The finite polygenic mixed model, Genetics 141 (1996) 1651–1656.
[29] Stricker C., Fernando R.L., Eltson R.C., An algorithm to approximate the likelihood for pedigree data with loops by cutting, Theor. Appl. Genet. 91 (1995) 1054–1063.
[30] Thomas A., Approximate computation of probability functions for pedigree analysis, IMJ J. Math. Appl. Med. Biol. 3 (1986) 157–166.
[31] Thomas D., Cortessis V., A Gibbs sampling approach to linkage analysis, Hum. Hered. 42 (1992) 63–76.
[32] Thompson E.A., Monte Carlo estimation of multilocus autozygosity probabilities, in: Sall J., Lehman A. (Eds.), Proceedings of the 1994 Interface Conference, Interface Foundation of North America: fairfax, Station, VA, 1994, pp. 498–506.
[33] Uimari P., Thaller G., Hoeschele I., The use of multiple markers in Bayesian methods for mapping quantitative trait loci, Genetics 143 (1996) 1831–1842.
[34] Wang T., Fernando R.L., Stricker C., Elston R.C., An approximation to the likelihood for a pedigree with loops, Theor. Appl. Genet. 93 (1996) 1299–1309.

APPENDIX

Proof that $\pi(\mathbf{m}) > 0$ implies $q(\mathbf{m}) > 0$

The proof is first given for a cut pedigree and then for a cut-extended pedigree.

Proof for cut pedigree

The probability of getting $\mathbf{m}$ from the target distribution is computed using equation (2). Even though the probability of getting $\mathbf{m}$ from the proposal distribution $q(\mathbf{m})$ was computed by multiplying probabilities that were used in

the sampling process, for this proof it is convenient to write it as

$$q(\mathbf{m}) \propto \prod_{j=1}^{n} q_j, \tag{A.1}$$

where

$$q_j = \begin{cases} \Pr(m_j) & \text{(a)} \\ \Pr(m_j|m_{m_j}, m_{f_j}) & \text{(b)} \\ \sum_{m_{m_j^*}} \Pr(m_j|m_{m_j^*}, m_{f_j}) \Pr(m_{m_j^*}) & \text{(c)} \\ \sum_{m_{f_j^*}} \Pr(m_j|m_{m_j}, m_{f_j^*}) \Pr(m_{f_j^*}) & \text{(d)} \\ \sum_{m_{j^*}} \Pr(m_{j^*}|m_{m_{j^*}}, m_{f_{j^*}}) & \text{(e)} \end{cases} \tag{A.2}$$

(a) if j is either a non-introduced or introduced founder and has known genotype
(b) if j is either a non-introduced or introduced offspring of either non-introduced or introduced parents, all of them with known genotypes
(c) if the mother of individual j is an introduced individual
(d) if the father of individual j is an introduced individual
(e) if j is an introduced individual.

Recall that $q(\mathbf{m})$ is the probability of sampling $\mathbf{m}$ from the cut pedigree conditional on $\mathbf{y}$, where $\mathbf{m}$ is the vector of missing genotypes of the original individuals and $\mathbf{y}$ is the vector of observed genotypes of the original and introduced individuals. Note that, in (A.2) the summations are over the missing genotypes of the introduced individuals. Also, note that for (a) and (b) $q_j = \pi_j$, when j is a non-introduced founder or a non-introduced offspring of non-introduced individuals. Furthermore, if j is an introduced founder or an offspring (introduced or non-introduced) of introduced parents, where all of them have known genotype, then $q_j > 0$. For (c), (d) and (e) individual j is a non-founder, therefore

$$\pi_j = \Pr(m_j|m_{m_j}, m_{f_j}).$$

As shown below, for (c), $\pi_j = \Pr(m_j|m_{m_j}, m_{f_j}) > 0$ implies $q_j > 0$. First, $\Pr(m_{m_{j^*}}) > 0$ for all $m_{m_{j^*}}$, and second, when $m_{m_{j^*}} = m_{m_j}$, $\Pr(m_j|m_{m_{j^*}}, m_{f_j}) = \pi_j$. Thus the term in (c) that corresponds to $m_{m_{j^*}} = m_{m_j}$ is greater than zero. Further, the other terms in (c) are greater than or equal to zero. So clearly, $q_j > 0$. Similarly, also for (d), $\pi_j = \Pr(m_j|m_{m_j}, m_{f_j}) > 0$ implies $q_j > 0$. Also for (e), as shown below, $\pi_j = \Pr(m_j|m_{m_j}, m_{f_j}) > 0$ implies $q_j > 0$. The term in (e) that corresponds to $m_{j^*} = m_j$ is π_j, which is greater than zero. The other terms in (e) are greater than or equal to zero, and so $q_j > 0$. From (2), $\pi(\mathbf{m}) > 0$ implies that $\pi_j > 0$ for all j. Further, as shown above, for (a) and (b) $q_j = \pi_j$, and for (c), (d) and (e) $\pi_j > 0$ implies $q_j > 0$. So, from (A.1), $q(\mathbf{m}) > 0$.

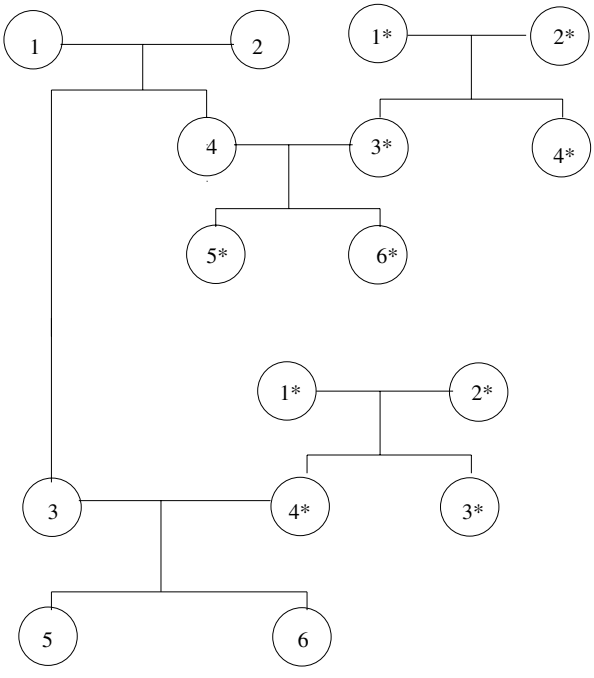

Figure 3. Cut-extended pedigree.

Proof for cut-extended pedigree

Consider the cut-extended pedigree in Figure 3. For this pedigree $q(\mathbf{m})$ can be written as

$$\begin{aligned} q(\mathbf{m}) \propto & \sum_{m_{3^*}} \sum_{m_{4^*}} \Pr(m_1) \Pr(m_2) \Pr(m_3|m_1, m_2) \Pr(m_4|m_1, m_2) \\ & \times \Pr(m_{1^*}) \Pr(m_{2^*}) \Pr(m_{3^*}|m_{1^*}, m_{2^*}) \Pr(m_{5^*}|m_{3^*}, m_4) \Pr(m_{6^*}|m_{3^*}, m_4) \\ & \times \Pr(m_{1^*}) \Pr(m_{2^*}) \Pr(m_{3^*}|m_{1^*}, m_{2^*}) \Pr(m_5|m_3, m_{4^*}) \Pr(m_6|m_3, m_{4^*}), \end{aligned} \tag{A.3}$$

where the summations are over the missing genotypes of the introduced individuals 3^* and 4^*. For one of the terms in (A.3) $m_{3^*} = m_3$ and $m_{4^*} = m_4$. In this term all of the factors are from $\pi(\mathbf{m})$, and therefore it will be greater than zero whenever $\pi(\mathbf{m}) > 0$. In general, $q(\mathbf{m})$ can be written as a sum of factors, and for one of its terms the missing genotype for each i^* will be equal to the sampled genotype for i. In this term all of the factors are from $\pi(\mathbf{m})$, and therefore it will be greater than zero whenever $\pi(\mathbf{m}) > 0$. The other terms in $q(\mathbf{m})$ are non-negative, therefore $q(\mathbf{m}) > 0$ whenever $\pi(\mathbf{m}) > 0$.