File S1

Details on the componentwise Markov chain Monte Carlo algorithm

Here we provide the computational details for the componentwise Markov chain Monte Carlo updates. Our aim is to sample from the joint posterior distribution of $f(\mathbf{M}, \pi, \kappa, \sigma, \delta, \lambda|\mathbf{n})$, which is specified by equation (4) and by the directed acyclic graph (DAG) in Figure 1. To do so, we use a combination of the Metropolis–Hastings algorithm and the Gibbs sampler for generating observations from $f(\mathbf{M}, \pi, \kappa, \sigma, \delta, \lambda|\mathbf{n})$ using outputs from a Markov chain (see, e.g., Gelman *et al.* 2004).

Each Markov chain is initialized with random values of the parameters drawn from their prior densities, except for the parameters p_{ij}, for which the observed frequencies are used, and the parameters π_js, for which the Laplace values are calculated from the dataset frequencies. The updating sequence is as follows: (*i*) all Ln_d parameters p_{ij}; (*ii*) all n_d parameters M_i; (*iii*) all L parameters π_j; (*iv*) the hyperparameter λ; (*v*) all L hyperparameters δ_j; (*vi*) all Ln_d parameters σ_{ij}; (*vii*) all Ln_d parameters κ_{ij}. Since the full posterior distribution of the model can be decomposed as a product over loci and over populations (see equation 4), each update only requires the re-computation of the relevant terms of the distribution $f(\mathbf{M}, \pi, \kappa, \sigma, \delta, \lambda|\mathbf{n})$. This improves the computational efficiency of the algorithm considerably.

The confluent hypergeometric, or Kummer's, functions ${}_1F_1(a; b; z)$ (see, e.g., Abramowitz and Stegun 1965, p. 504) were computed following a procedure proposed by Pearson (Pearson 2009), which is based on the power series definition of the function:

$$ {}_1F_1(a; b; z) = \sum_{j=0}^{\infty} \underbrace{\frac{(a)_j}{(b)_j} \frac{z^j}{j!}}_{A_j}, \tag{S1.1}$$

where, for some parameter p, the Pochhammer symbol $(p)_j$ is defined as:

$$ (p)_0 = 1, \quad (p)_j = p(p+1) \dots (p+j-1), \quad \text{for } j = 1, 2, \dots \;. \tag{S1.2}$$

The computation of the terms of the power series in equation (S1.1) can then be carried out using the following procedure:

$$A_0 = S_0 = 1,$$
$$A_{j+1} = A_j \frac{a+j}{b+j}\frac{z}{j+1}, \tag{S1.3}$$
$$S_{j+1} = S_j + A_{j+1}, \quad \text{for } j = 1, 2, \ldots$$

where A_j represents the $(j+1)$th term of the power series in equation (S1.1), and S_j represents the sum of the first $(j+1)$ terms. The computation was stopped when both $|A_N|/|S_{N-1}| < 10^{-12}$ and $|A_{N+1}|/|S_N| < 10^{-12}$. This criterion is equivalent to truncating the series in equation (S1.1), and requires that two consecutive terms to be small compared to the sum already computed.

Updating p_{ij}: The parameters p_{ij} are updated iteratively in each deme, one locus at a time. In the ith deme, at locus j, one allele is chosen at random from a Bernoulli trial with probability 0.5. The new allele frequency p'_{ij} is chosen as a random variable drawn from a uniform distribution around the current value p_{ij}:

$$p'_{ij} \sim U\left(p_{ij} - \Delta_p, p_{ij} + \Delta_p\right). \tag{S1.4}$$

The size of the interval Δ_p is a constant, which is adjusted during 25 short pilot runs of 1,000 iterations, in order to get acceptance rates between 0.25 and 0.40 (see, e.g., Gilks *et al.* 1996). Since p_{ij} is a frequency comprised between 0 and 1, if p'_{ij} is outside the interval $[0, 1]$, the excess is reflected back into the interval; that is, if $p'_{ij} < 0$ then p'_{ij} is reset to its absolute value $|p'_{ij}|$, and if $p'_{ij} > 1$ then p'_{ij} is reset to $2 - p'_{ij}$. This proposal is symmetric (Yang 2005). The updated allele frequency p'_{ij} is therefore accepted according to the appropriate Metropolis probability, which reads:

$$1 \wedge \frac{\mathcal{L}(p'_{ij}; \mathbf{n}_{ij})\psi(p'_{ij}; M_i, \pi_j, \kappa_{ij}, \sigma_{ij})}{\mathcal{L}(p_{ij}; \mathbf{n}_{ij})\psi(p_{ij}; M_i, \pi_j, \kappa_{ij}, \sigma_{ij})}. \tag{S1.5}$$

Equation (S1.5) can be rewritten as

$$1 \wedge \exp\left[\sigma_{ij}\left(\tilde{p}'_{ij} - \tilde{p}_{ij}\right)\right] \frac{p'^{x_{ij}+M_i\pi_j-1}_{ij}(1-p'_{ij})^{(n_{ij}-x_{ij})M_i+(1-\pi_j)-1}}{p^{x_{ij}+M_i\pi_j-1}_{ij}(1-p_{ij})^{(n_{ij}-x_{ij})M_i+(1-\pi_j)-1}}, \quad \text{(S1.6)}$$

where $\tilde{p}'_{ij} \equiv \kappa_{ij}(1-p'_{ij}) + (1-\kappa_{ij})p'_{ij}$.

Updating M_i: The parameters M_i are updated iteratively, one deme at a time. The proposed value M'_i is drawn from a lognormal distribution with median equal to the current value M_i, i.e.:

$$q(M_i \to M'_i) = \frac{1}{M'_i \nu_M \sqrt{2\pi}} \exp\left(\frac{-\ln(M'_i/M_i)^2}{2\nu_M^2}\right), \quad \text{(S1.7)}$$

where ν_M is the standard deviation on the log scale. The standard deviation ν_M is a constant, which is adjusted during 25 short pilot runs of 1,000 iterations, in order to get acceptance rates between 0.25 and 0.40. Because the lognormal jumping rule is asymmetric, a Metropolis–Hastings update is required in which the Metropolis ratio is weighted by the ratio of the forward and reverse proposal densities (which is sometimes referred to as the ''Hastings term'': see, e.g., Gelman *et al.* 2004, p. 291). This means that when some moves are more likely to happen (because of the asymmetry of the proposal distribution), their probability of acceptance is decreased proportionately. Here, the ratio $q(M'_i \to M_i)/q(M_i \to M'_i)$ reduces to M'_i/M_i. In order to avoid computational problems with excessively small or large M_i values, all moves falling outside the interval $[0.0011, 000]$ are discarded (i.e., the chain is kept unchanged). Otherwise, the proposed value M'_i is accepted according to the appropriate Metropolis–Hastings probability, which is:

$$1 \wedge \frac{\left[\prod_{j=1}^{L} \psi(p_{ij}; M'_i, \pi_j, \kappa_{ij}, \sigma_{ij})\right] f(M'_i) q(M'_i \to M_i)}{\left[\prod_{j=1}^{L} \psi(p_{ij}; M_i, \pi_j, \kappa_{ij}, \sigma_{ij})\right] f(M_i) q(M_i \to M'_i)}. \quad \text{(S1.8)}$$

Equation (S1.8) can be rewritten as

$$1\wedge\left[\frac{\Gamma(M_i)}{\Gamma(M'_i)}\right]^L \frac{\prod_{j=1}^{L}\Gamma(M_i\pi_j)\Gamma(M_i(1-\pi_j))_1F_1(M_i\tilde{\pi}_{ij};M_i;\sigma_{ij})p_{ij}^{M'_i\pi_j}(1-p_{ij})^{M'_i(1-\pi_j)}}{\prod_{j=1}^{L}\Gamma(M'_i\pi_j)\Gamma(M'_i(1-\pi_j))_1F_1(M'_i\tilde{\pi}_{ij};M'_i;\sigma_{ij})p_{ij}^{M_i\pi_j}(1-p_{ij})^{M_i(1-\pi_j)}} \tag{S1.9}$$

Updating π_j: The parameters π_j are updated iteratively, one locus at a time. In the ith deme, at locus j, one allele is chosen at random from a Bernoulli trial with probability 0.5. The proposed allele frequency π'_j is chosen as a random variable drawn from a uniform distribution around the current value π_j:

$$\pi'_j \sim U\left(\pi_j - \Delta_\pi, \pi_j + \Delta_\pi\right). \tag{S1.10}$$

The size of the interval Δ_π is a constant, which is adjusted during 25 short pilot runs of 1,000 iterations, in order to get acceptance rates between 0.25 and 0.40. Since π_j is a frequency comprised between 0 and 1, if π'_j is outside the interval $[0,1]$, the excess is reflected back into the interval; that is, if $\pi'_j < 0$ then π'_j is reset to its absolute value $|\pi'_j|$, and if $\pi'_j > 1$ then π'_j is reset to $2-\pi'_j$. This proposal is symmetric, and the updated allele frequency π'_j is therefore accepted according to the appropriate Metropolis probability, which reads:

$$1\wedge\frac{\left[\prod_{i=1}^{n_\mathrm{d}}\psi(p_{ij};M_i,\pi'_j,\kappa_{ij},\sigma_{ij})\right]f(\pi'_j)}{\left[\prod_{i=1}^{n_\mathrm{d}}\psi(p_{ij};M_i,\pi_j,\kappa_{ij},\sigma_{ij})\right]f(\pi_j)}. \tag{S1.11}$$

Equation (S1.11) can be rewritten as

$$1\wedge\frac{\prod_{i=1}^{n_\mathrm{d}}\Gamma(M_i\pi_j)\Gamma(M_i(1-\pi_j))_1F_1(M_i\tilde{\pi}_{ij};M_i;\sigma_{ij})p_{ij}^{M_i\pi'_j}(1-p_{ij})^{M_i(1-\pi'_j)}}{\prod_{i=1}^{n_\mathrm{d}}\Gamma(M_i\pi'_j)\Gamma(M_i(1-\pi'_j))_1F_1(M_i\tilde{\pi}'_{ij};M_i;\sigma_{ij})p_{ij}^{M_i\pi_j}(1-p_{ij})^{M_i(1-\pi_j)}}, \tag{S1.12}$$

where $\tilde{\pi}'_{ij} \equiv \kappa_{ij}(1-\pi'_j)+(1-\kappa_{ij})\pi'_j$.

Updating λ: The proposed value of the hyperparameter λ' is drawn from a lognormal distribution with median equal to the current value λ, i.e.:

$$q(\lambda \to \lambda') = \frac{1}{\lambda' \nu_\lambda \sqrt{2\pi}} \exp\left(\frac{-\ln(\lambda'/\lambda)^2}{2\nu_\lambda^2} \right), \tag{S1.13}$$

where ν_λ is the standard deviation on the log scale. The standard deviation ν_λ is a constant, which is adjusted during 25 short pilot runs of 1,000 iterations, in order to get acceptance rates between 0.25 and 0.40. Because the lognormal jumping rule is asymmetric, a Metropolis–Hastings update is required in which the Metropolis ratio is weighted by the ratio of the forward and reverse proposal densities. This means that when some moves are more likely to happen (because of the asymmetry of the proposal distribution), their probability of acceptance is decreased proportionately. Here, the ratio $q(\lambda' \to \lambda)/q(\lambda \to \lambda')$ reduces to λ'/λ. In order to avoid computational problems with excessively small or large λ' values, all moves falling outside the interval $[0, 500]$ are discarded (i.e., the chain is kept unchanged). Otherwise, the proposed value λ' is accepted according to the appropriate Metropolis–Hastings probability, which is:

$$1 \wedge \frac{\left[\prod_{j=1}^{L} f(\delta_j|\lambda')\right] f(\lambda'|\Lambda) q(\lambda' \to \lambda)}{\left[\prod_{j=1}^{L} f(\delta_j|\lambda)\right] f(\lambda|\Lambda) q(\lambda \to \lambda')}. \tag{S1.14}$$

Equation (S1.14) can be rewritten as

$$1 \wedge \left(\frac{\lambda}{\lambda'}\right)^{L-1} \exp\left[(\lambda' - \lambda) \left(\frac{\sum_{j=1}^{L} \delta_j}{\lambda\lambda'} - \frac{1}{\Lambda} \right) \right] \tag{S1.15}$$

Updating δ_j: The parameters δ_j are updated iteratively, one locus at a time. The proposed value of the hyperparameters δ'_j is drawn from a lognormal distribution with median equal to the current value δ_j, i.e.:

$$q(\delta_j \to \delta'_j) = \frac{1}{\delta'_j \nu_\delta \sqrt{2\pi}} \exp\left(\frac{-\ln(\delta'_j/\delta_j)^2}{2\nu_\delta^2} \right), \tag{S1.16}$$

where ν_δ is the standard deviation on the log scale. The standard deviation ν_δ is a constant, which is adjusted during 25 short pilot runs of 1,000 iterations, in order to get acceptance rates between 0.25 and 0.40. Because the lognormal jumping rule is asymmetric, a Metropolis–Hastings update is required in which the Metropolis ratio is weighted by the ratio of the forward and reverse proposal densities. This means that when some moves are more likely to happen (because of the asymmetry of the proposal distribution), their probability of acceptance is decreased proportionately. Here, the ratio $q(\delta'_j \to \delta_j)/q(\delta_j \to \delta'_j)$ reduces to δ'_j/δ_j. In order to avoid computational problems with excessively small or large δ_j values, all moves falling outside the interval $[0, 500]$ are discarded (i.e., the chain is kept unchanged). Otherwise, the proposed value δ'_j is accepted according to the appropriate Metropolis–Hastings probability, which is:

$$1 \wedge \frac{\left[\prod_{i=1}^{n_\mathrm{d}} f(\sigma_{ij}|\delta'_j)\right] f(\delta'_j|\lambda) q(\delta'_j \to \delta_j)}{\left[\prod_{i=1}^{n_\mathrm{d}} f(\sigma_{ij}|\delta_j)\right] f(\delta_j|\lambda) q(\delta_j \to \delta'_j)}. \tag{S1.17}$$

Equation (S1.17) can be rewritten as

$$1 \wedge \left(\frac{\delta_j}{\delta'_j}\right)^{n_\mathrm{d}-1} \exp\left[\left(\delta'_j - \delta_j\right)\left(\frac{\sum_{i=1}^{n_\mathrm{d}} \sigma_{ij}}{\delta_j \delta'_j} - \frac{1}{\lambda}\right)\right] \tag{S1.18}$$

Updating σ_{ij}: The parameters σ_{ij} are updated iteratively in each deme, one locus at a time. In the ith deme, at locus j, the proposed value of the parameters σ'_{ij} is drawn from a lognormal distribution with median equal to the current value σ_{ij}, i.e.:

$$q(\sigma_{ij} \to \sigma'_{ij}) = \frac{1}{\sigma'_{ij}\nu_\sigma\sqrt{2\pi}} \exp\left(\frac{-\ln(\sigma'_{ij}/\sigma_{ij})^2}{2\nu_\sigma^2}\right), \tag{S1.19}$$

where ν_σ is the standard deviation on the log scale. The standard deviation ν_σ is a constant, which is adjusted during 25 short pilot runs of 1,000 iterations, in order to get acceptance rates between 0.25 and 0.40. Because the lognormal jumping rule is asymmetric, a Metropolis–Hastings update is required in which the Metropolis ratio is weighted by the ratio of the forward and reverse proposal densities. This means that when some moves are more likely to happen (because of the asymmetry of the proposal distribution), their probability of acceptance is decreased proportionately. Here, the ratio $q(\sigma'_{ij} \to \sigma_{ij})/q(\sigma_{ij} \to \sigma'_{ij})$ reduces to σ'_{ij}/σ_{ij}. In order to avoid computational problems with excessively small or large σ_{ij} values, all moves falling outside the interval $[0, 500]$ are discarded (i.e., the chain is kept unchanged). Otherwise, the proposed value σ'_{ij} is accepted according to the appropriate Metropolis–Hastings probability, which is:

$$\frac{\psi(p_{ij}; M_i, \pi_j, \kappa_{ij}, \sigma'_{ij}) f(\sigma'_{ij}|\delta_j) q(\sigma'_{ij} \to \sigma_{ij})}{\psi(p_{ij}; M_i, \pi_j, \kappa_{ij}, \sigma_{ij}) f(\sigma_{ij}|\delta_j) q(\sigma_{ij} \to \sigma'_{ij})}. \tag{S1.20}$$

Equation (S1.20) can be rewritten as

$$\frac{\sigma'_{ij}}{\sigma_{ij}} \exp\left[(\sigma'_{ij} - \sigma_{ij})\left(\tilde{p}_{ij} - \frac{1}{\delta_j}\right)\right] \frac{{}_1F_1(M_i\tilde{\pi}_{ij}; M_i; \sigma_{ij})}{{}_1F_1(M_i\tilde{\pi}_{ij}; M_i; \sigma'_{ij})}. \tag{S1.21}$$

Updating κ_{ij}: The parameters κ_{ij} are updated iteratively in each deme, one locus at a time. In the ith deme, at locus j, the variable κ_{ij}, which indicates which of the two alleles is selected for, is updated using Gibbs sampling based on the conditional posterior distribution:

$$f(\kappa_{ij}|\theta_{[-\kappa_{ij}]}) \propto \psi(p_{ij}; M_i, \pi_j, \kappa_{ij}, \sigma_{ij}) f(\kappa_{ij}), \tag{S1.22}$$

where $\theta_{[-\kappa_{ij}]}$ represents all the model parameters but κ_{ij}. Since κ_{ij} can only take two integer values (0 and 1), it can be shown that:

$$\Pr(\kappa_{ij} = 0|\theta_{[-\kappa_{ij}]}) \propto \frac{1}{2}\left[\frac{\exp\left[\sigma_{ij}p_{ij}\right]}{{}_1F_1(M_i\pi_j; M_i; \sigma_{ij})}\right], \tag{S1.23}$$

and

$$\Pr(\kappa_{ij} = 1|\theta_{[-\kappa_{ij}]}) \propto \frac{1}{2}\left[\frac{\exp\left[\sigma_{ij}(1 - p_{ij})\right]}{{}_1F_1(M_i(1 - \pi_j); M_i; \sigma_{ij})}\right]. \tag{S1.24}$$

Therefore, the conditional posterior distribution of $\left(\kappa_{ij}|\theta_{[-\kappa_{ij}]}\right)$ from equation (S1.22) can be rewritten as

$$\left(\kappa_{ij}|\theta_{[-\kappa_{ij}]}\right) \sim \mathrm{Bernoulli}\left(\rho\right), \tag{S1.25}$$

where

$$\begin{aligned} \rho &\equiv \frac{\Pr(\kappa_{ij}=0|\theta_{[-\kappa_{ij}]})}{\Pr(\kappa_{ij}=0|\theta_{[-\kappa_{ij}]}) + \Pr(\kappa_{ij}=1|\theta_{[-\kappa_{ij}]})} \\ &= \left[1 + \frac{{}_1F_1(M_i\pi_{ij}; M_i; \sigma_{ij})}{{}_1F_1(M_i(1-\pi_{ij}); M_i; \sigma_{ij})} \exp\left[\sigma_{ij}(1-2p_{ij})\right]\right]^{-1}. \end{aligned} \tag{S1.26}$$

Literature Cited

Abramowitz, M., and I. A. Stegun, 1965 *Handbook of Mathematical Functions*. Dover Publication, Inc., New York.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2004 *Bayesian Data Analysis*. Chapman & Hall, New York, 2nd edition.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, 1996 *Markov Chain Monte Carlo in Practice*. Chapman & Hall, New York, 2nd edition.

Pearson, J., 2009 *Computation of Hypergeometric Functions*. Ph.D. thesis, University of Oxford.

Yang, Z., 2005 Bayesian inference in molecular phylogenetics. In O. Gascuel, editor, *Mathematics of Evolution and Phylogeny*. Oxford University Press, Oxford, 63–90.