Stochastic Variational Inference in Phylogenetics

Tung Dang,[*,1] *and Hirohisa Kishino*[1]
[1] Laboratory of Biometrics and Bioinformatics, University of Tokyo, Tokyo, Japan

A. Variational Inference for JC69 model

The D data are two aligned sequences, each n sites long, with x difference. The likelihood is given by the JC69 model as:

$$p\left(D|d\right) = \left(\frac{1}{4}p_1\right)^x \left(\frac{1}{4}p_0\right)^{n-x} = \left(\frac{1}{16} - \frac{1}{16}e^{\frac{-4d}{3}}\right)^x \left(\frac{1}{16} + \frac{3}{16}e^{\frac{-4d}{3}}\right)^{n-x} \tag{1}$$

In Variational inference framework, the optimal solution to the posterior distribution of d is

$$\begin{aligned}
&\log q^*\left(d|\gamma,\lambda\right) = \mathrm{E}_q\left[\log p\left(d\right) + \log\left(D|d\right)\right] \\
&= \mathrm{E}_q\left[x\log\left(\frac{1}{16} - \frac{1}{16}e^{\frac{-4d}{3}}\right)\right] + \mathrm{E}_q\left[\left(n-x\right)\log\left(\frac{1}{16} + \frac{3}{16}e^{\frac{-4d}{3}}\right)\right] \\
&+\mathrm{E}_q\left[\log\left(\frac{(\beta)^\alpha}{\Gamma(\alpha)}\right) + (\alpha-1)\log\left(d\right) - \beta d\right]
\end{aligned} \tag{2}$$

We consider second-order Taylor expansion for $\log\left(\frac{1}{16} - \frac{1}{16}e^{\frac{-4d}{3}}\right)$ and $\log\left(\frac{1}{16} + \frac{3}{16}e^{\frac{-4d}{3}}\right)$ for d at $d^{'} = \frac{\gamma}{\lambda}$

$$\begin{aligned}
&\log\left(\frac{1}{16} - \frac{1}{16}e^{\frac{-4d}{3}}\right) \approx \log\left(\frac{1}{16} - \frac{1}{16}e^{\frac{-4d^{'}}{3}}\right) + \frac{\partial\log\left(\frac{1}{16} - \frac{1}{16}e^{\frac{-4d}{3}}\right)}{\partial d}\left(d-d^{'}\right) \\
&+\left(d-d^{'}\right)\frac{\partial^2\log\left(\frac{1}{16} - \frac{1}{16}e^{\frac{-4d}{3}}\right)}{\partial d^2}\left(d-d^{'}\right)
\end{aligned} \tag{3}$$

$$\begin{aligned}
&\frac{\partial\log\left(\frac{1}{16} - \frac{1}{16}e^{\frac{-4d}{3}}\right)}{\partial d}|_{d=d^{'}} = \frac{1}{\frac{1}{16} - \frac{1}{16}e^{\frac{-4d^{'}}{3}}}\frac{\partial\left(\frac{1}{16} - \frac{1}{16}e^{\frac{-4d}{3}}\right)}{\partial d}|_{d=d^{'}} \\
&= \frac{e^{\frac{-4d^{'}}{3}}}{12\left(\frac{1}{16} - \frac{1}{16}e^{\frac{-4d^{'}}{3}}\right)} = \frac{4}{3e^{\frac{4d^{'}}{3}} - 3}
\end{aligned} \tag{4}$$

$$\frac{\partial^2\log\left(\frac{1}{16} - \frac{1}{16}e^{\frac{-4d^{'}}{3}}\right)}{\partial d^2}|_{d=d^{'}} = 4\frac{\partial}{\partial d}\left(\frac{1}{3e^{\frac{4d}{3}} - 3}\right) = -\frac{16e^{\frac{4d^{'}}{3}}}{\left(3e^{\frac{4d^{'}}{3}} - 3\right)^2} \tag{5}$$

$$\log\left(\frac{1}{16}+\frac{3}{16}e^{\frac{-4d}{3}}\right)\approx\log\left(\frac{1}{16}+\frac{3}{16}e^{\frac{-4d'}{3}}\right)+\frac{\partial\log\left(\frac{1}{16}+\frac{3}{16}e^{\frac{-4d}{3}}\right)}{\partial d}\left(d-d'\right)$$
$$+\left(d-d'\right)\frac{\partial^2\log\left(\frac{1}{16}+\frac{3}{16}e^{\frac{-4d}{3}}\right)}{\partial d^2}\left(d-d'\right) \tag{6}$$

$$\frac{\partial\log\left(\frac{1}{16}+\frac{3}{16}e^{\frac{-4d}{3}}\right)}{\partial d}|_{d=d'}=\frac{1}{\frac{1}{16}+\frac{3}{16}e^{\frac{-4d'}{3}}}\frac{\partial\left(\frac{1}{16}+\frac{3}{16}e^{\frac{-4d}{3}}\right)}{\partial d}|_{d=d'}$$
$$=-\frac{e^{\frac{-4d'}{3}}}{4\left(\frac{1}{16}+\frac{3}{16}e^{\frac{-4d'}{3}}\right)}=-\frac{4}{e^{\frac{4d'}{3}}+3} \tag{7}$$

$$\frac{\partial^2\log\left(\frac{1}{16}+\frac{3}{16}e^{\frac{-4d}{3}}\right)}{\partial d^2}|_{d=d'}=-4\frac{\partial}{\partial d}\left(\frac{1}{e^{\frac{4d}{3}}+3}\right)=\frac{16e^{\frac{4d'}{3}}}{3\left(e^{\frac{4d'}{3}}+3\right)^2} \tag{8}$$

With result of first-order Taylor expansion and the principle of the VI framework, the optimal solution to the posterior distribution of d is

$$\log q^*\left(d|\gamma,\lambda\right)\approx x\mathrm{E}_q\left[\log\left(\frac{1}{16}-\frac{1}{16}e^{\frac{-4d'}{3}}\right)+\left(\frac{4}{3e^{\frac{4d'}{3}}-3}\right)\left(d-d'\right)\right]$$
$$+\left(n-x\right)\mathrm{E}_q\left[\log\left(\frac{1}{16}+\frac{3}{16}e^{\frac{-4d'}{3}}\right)+\left(-\frac{4}{e^{\frac{4d'}{3}}+3}\right)\left(d-d'\right)\right] \tag{9}$$
$$+\mathrm{E}_q\left[\log\left(\frac{(\beta)^\alpha}{\Gamma(\alpha)}\right)+(\alpha-1)\log\left(d\right)-\beta d\right]$$

$$\log q^*\left(d|\gamma,\lambda\right)\approx x\mathrm{E}_q\left[\left(\frac{4}{3e^{\frac{4d'}{3}}-3}\right)d\right]+\left(n-x\right)\mathrm{E}_q\left[\left(-\frac{4}{e^{\frac{4d'}{3}}+3}\right)d\right] \tag{10}$$
$$+\mathrm{E}_q\left[(\alpha-1)\log\left(d\right)-\beta d\right]+const$$

which has the logarithm form of the gamma distribution.

B. Variational Inference for K80 model

Here we illustrate the major features of Variational Inference by applying it to the problem of estimating d and the transition/transversion rate ratio κ under the K80 model using a pair of DNA sequences. D is an alignment of the human and orangutan mitochondrial 12S rRNA genes, summarized as $n_S = 84$ transitional differences and $n_V = 6$ transversional differences at n = 948 sites. We assign independent gamma priors,

$$p(d) = Gamma\left(d|\alpha_d, \beta_d\right) = \frac{(\beta_d)^{\alpha_d}}{\Gamma(\alpha_d)} d^{\alpha_d - 1} e^{-\beta_d d} with \alpha_d = 2, \beta_d = 20$$
$$p(\kappa) = Gamma\left(\kappa|\alpha_\kappa, \beta_\kappa\right) = \frac{(\beta_\kappa)^{\alpha_\kappa}}{\Gamma(\alpha_\kappa)} \kappa^{\alpha_\kappa - 1} e^{-\beta_\kappa \kappa} with \alpha_\kappa = 2, \beta_\kappa = 0.1$$

The likelihood is given by the K80 model as:

$$p\left(D|d,\kappa\right) = \left(\frac{1}{4}p_0\right)^{n-n_S-n_V} \left(\frac{1}{4}p_1\right)^{n_S} \left(\frac{1}{4}p_2\right)^{n_V}$$

$$p_0 = \frac{1}{4} + \frac{1}{4}e^{-\frac{4d}{\kappa+2}} + \frac{1}{2}e^{-2d\frac{\kappa+1}{\kappa+2}}$$
$$p_1 = \frac{1}{4} + \frac{1}{4}e^{-\frac{4d}{\kappa+2}} - \frac{1}{2}e^{-2d\frac{\kappa+1}{\kappa+2}}$$
$$p_2 = \frac{1}{4} - \frac{1}{4}e^{-\frac{4d}{\kappa+2}}$$

In Variational inference framework, the optimal solution to the posterior distribution of d is

$$\begin{aligned}
&\log q^*\left(d|\gamma,\gamma^{'}\right) = \mathrm{E}_q\left[\log p\left(d\right) + \log\left(D|d,\kappa\right)\right]\\
&= \mathrm{E}_q\left[\left(n - n_S - n_V\right)\log\left(\frac{1}{16} + \frac{1}{16}e^{-\frac{4d}{\kappa+2}} + \frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)\right]\\
&+\mathrm{E}_q\left[\left(n_S\right)\log\left(\frac{1}{16} + \frac{1}{16}e^{-\frac{4d}{\kappa+2}} - \frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)\right]\\
&+\mathrm{E}_q\left[\left(n_V\right)\log\left(\frac{1}{16} - \frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)\right]\\
&+\mathrm{E}_q\left[\log\left(\frac{(\beta_d)^{\alpha_d}}{\Gamma(\alpha_d)}\right) + \left(\alpha_d - 1\right)\log\left(d\right) - \beta_d d\right]
\end{aligned}$$

$$\begin{aligned}
&\log q^*\left(\kappa|\lambda,\lambda^{'}\right) = \mathrm{E}_q\left[\log p\left(\kappa\right) + \log\left(D|d,\kappa\right)\right]\\
&= \mathrm{E}_q\left[\left(n - n_S - n_V\right)\log\left(\frac{1}{16} + \frac{1}{16}e^{-\frac{4d}{\kappa+2}} + \frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)\right]\\
&+\mathrm{E}_q\left[\left(n_S\right)\log\left(\frac{1}{16} + \frac{1}{16}e^{-\frac{4d}{\kappa+2}} - \frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)\right]\\
&+\mathrm{E}_q\left[\left(n_V\right)\log\left(\frac{1}{16} - \frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)\right]\\
&+\mathrm{E}_q\left[\log\left(\frac{(\beta_\kappa)^{\alpha_\kappa}}{\Gamma(\alpha_\kappa)}\right) + \left(\alpha_\kappa - 1\right)\log\left(\kappa\right) - \beta_\kappa \kappa\right]
\end{aligned}$$

We consider second-order Taylor expansion for $\log\left(\frac{1}{16} + \frac{1}{16}e^{-\frac{4d}{\kappa+2}} + \frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)$ and $\log\left(\frac{1}{16} + \frac{1}{16}e^{-\frac{4d}{\kappa+2}} - \frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)$ and $\log\left(\frac{1}{16} - \frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)$ for d and κ at $d^{'} = \frac{\gamma}{\gamma^{'}}$ and $\kappa^{'} = \frac{\lambda}{\lambda^{'}}$

For $\log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}+\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)$

$$\log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}+\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)\approx\log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d^{'}}{\kappa^{'}+2}}+\frac{1}{8}e^{-2d^{'}\frac{\kappa^{'}+1}{\kappa^{'}+2}}\right)$$
$$+\frac{\partial\log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}+\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial d}\left(d-d^{'}\right)+\frac{\partial\log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}+\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial\kappa}\left(\kappa-\kappa^{'}\right)$$
$$+\frac{\partial^2\log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}+\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial d\partial\kappa}\left(d-d^{'}\right)\left(\kappa-\kappa^{'}\right)$$
$$+\frac{1}{2}\left(d-d^{'}\right)\frac{\partial^2\log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}+\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial d^2}\left(d-d^{'}\right)$$
$$+\frac{1}{2}\left(\kappa-\kappa^{'}\right)\frac{\partial^2\log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}+\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial\kappa^2}\left(\kappa-\kappa^{'}\right)$$

$$\frac{\partial\log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}+\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial d}|_{d=d^{'},\kappa=\kappa^{'}}$$
$$=\frac{1}{\frac{1}{16}+\frac{1}{16}e^{-\frac{4d^{'}}{\kappa^{'}+2}}+\frac{1}{8}e^{-2d^{'}\frac{\kappa^{'}+1}{\kappa^{'}+2}}}\frac{\partial\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}+\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial d}|_{d=d^{'},\kappa=\kappa^{'}}$$
$$=\frac{1}{\frac{1}{16}+\frac{1}{16}e^{-\frac{4d^{'}}{\kappa^{'}+2}}+\frac{1}{8}e^{-2d^{'}\frac{\kappa^{'}+1}{\kappa^{'}+2}}}\left(-\frac{\left(\kappa^{'}+1\right)e^{-2d^{'}\frac{\kappa^{'}+1}{\kappa^{'}+2}}}{4\left(\kappa^{'}+2\right)}-\frac{e^{-\frac{4d^{'}}{\kappa^{'}+2}}}{4\left(\kappa^{'}+2\right)}\right)$$
$$=-\frac{4\left(e^{\frac{2d^{'}\left(\kappa^{'}+1\right)}{\kappa^{'}+2}}+\left(\kappa^{'}+1\right)e^{\frac{4d^{'}}{(\kappa^{'}+2)}}\right)}{(\kappa^{'}+2)\left(\left(e^{\frac{4d^{'}}{(\kappa^{'}+2)}}+1\right)e^{\frac{2d^{'}\left(\kappa^{'}+1\right)}{\kappa^{'}+2}}+2e^{\frac{4d^{'}}{(\kappa^{'}+2)}}\right)}$$

$$\frac{\partial \log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}+\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial \kappa}|_{d=d^{'},\kappa=\kappa^{'}}$$

$$=\frac{1}{\frac{1}{16}+\frac{1}{16}e^{-\frac{4d^{'}}{\kappa^{'}+2}}+\frac{1}{8}e^{-2d^{'}\frac{\kappa^{'}+1}{\kappa^{'}+2}}}\frac{\partial\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}+\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial \kappa}|_{d=d^{'},\kappa=\kappa^{'}}$$

$$=\frac{1}{\frac{1}{16}+\frac{1}{16}e^{-\frac{4d^{'}}{\kappa^{'}+2}}+\frac{1}{8}e^{-2d^{'}\frac{\kappa^{'}+1}{\kappa^{'}+2}}}\left(\frac{d^{'}e^{-\frac{4d^{'}}{\left(\kappa^{'}+2\right)}}}{4\left(\kappa^{'}+2\right)^{2}}-\frac{d^{'}e^{-\frac{2d^{'}\left(\kappa^{'}+1\right)}{\kappa^{'}+2}}}{4\left(\kappa^{'}+2\right)^{2}}\right)$$

$$=\frac{4d^{'}\left(e^{\frac{2d^{'}\left(\kappa^{'}+1\right)}{\kappa^{'}+2}}-e^{\frac{4d^{'}}{\kappa^{'}+2}}\right)}{\left(\kappa^{'}+2\right)^{2}\left(\left(e^{\frac{4d^{'}}{\left(\kappa^{'}+2\right)}}+1\right)e^{\frac{2d^{'}\left(\kappa^{'}+1\right)}{\kappa^{'}+2}}+2e^{\frac{4d^{'}}{\left(\kappa^{'}+2\right)}}\right)}$$

For $\log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}-\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)$

$$\log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}-\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)\approx\log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d^{'}}{\kappa^{'}+2}}-\frac{1}{8}e^{-2d^{'}\frac{\kappa^{'}+1}{\kappa^{'}+2}}\right)$$
$$+\frac{\partial \log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}-\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial d}\left(d-d^{'}\right)+\frac{\partial \log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}-\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial \kappa}\left(\kappa-\kappa^{'}\right)$$
$$+\frac{\partial^2 \log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}-\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial d\partial \kappa}\left(d-d^{'}\right)\left(\kappa-\kappa^{'}\right)$$
$$+\frac{1}{2}\left(d-d^{'}\right)\frac{\partial^2 \log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}-\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial d^2}\left(d-d^{'}\right)$$
$$+\frac{1}{2}\left(\kappa-\kappa^{'}\right)\frac{\partial^2 \log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}-\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial \kappa^2}\left(\kappa-\kappa^{'}\right)$$

$$\frac{\partial \log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}-\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial d}|_{d=d',\kappa=\kappa'}$$

$$=\frac{1}{\frac{1}{16}+\frac{1}{16}e^{-\frac{4d'}{\kappa'+2}}-\frac{1}{8}e^{-2d'\frac{\kappa'+1}{\kappa'+2}}}\frac{\partial\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}-\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial d}|_{d=d',\kappa=\kappa'}$$

$$=\frac{1}{\frac{1}{16}+\frac{1}{16}e^{-\frac{4d'}{\kappa'+2}}-\frac{1}{8}e^{-2d'\frac{\kappa'+1}{\kappa'+2}}}\left(\frac{\left(\kappa'+1\right)e^{-2d'\frac{\kappa'+1}{\kappa'+2}}}{4\left(\kappa'+2\right)}-\frac{e^{-\frac{4d'}{\kappa'+2}}}{4\left(\kappa'+2\right)}\right)$$

$$=\frac{4\left(e^{\frac{2d'\left(\kappa'+1\right)}{\kappa'+2}}+\left(-\kappa'-1\right)e^{\frac{4d'}{(\kappa'+2)}}\right)}{(\kappa'+2)\left(\left(e^{\frac{4d'}{(\kappa'+2)}}+1\right)e^{\frac{2d'\left(\kappa'+1\right)}{\kappa'+2}}+2e^{\frac{4d'}{(\kappa'+2)}}\right)}$$

$$\frac{\partial \log\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}-\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial \kappa}|_{d=d',\kappa=\kappa'}$$

$$=\frac{1}{\frac{1}{16}+\frac{1}{16}e^{-\frac{4d'}{\kappa'+2}}-\frac{1}{8}e^{-2d'\frac{\kappa'+1}{\kappa'+2}}}\frac{\partial\left(\frac{1}{16}+\frac{1}{16}e^{-\frac{4d}{\kappa+2}}-\frac{1}{8}e^{-2d\frac{\kappa+1}{\kappa+2}}\right)}{\partial \kappa}|_{d=d',\kappa=\kappa'}$$

$$=\frac{1}{\frac{1}{16}+\frac{1}{16}e^{-\frac{4d'}{\kappa'+2}}-\frac{1}{8}e^{-2d'\frac{\kappa'+1}{\kappa'+2}}}\left(\frac{d'e^{-\frac{4d'}{(\kappa'+2)}}}{4\left(\kappa'+2\right)^2}+\frac{d'e^{-\frac{2d'\left(\kappa'+1\right)}{\kappa'+2}}}{4\left(\kappa'+2\right)^2}\right)$$

$$=\frac{4d'\left(e^{\frac{2d'\left(\kappa'+1\right)}{\kappa'+2}}+e^{\frac{4d'}{\kappa'+2}}\right)}{(\kappa'+2)^2\left(\left(e^{\frac{4d'}{(\kappa'+2)}}+1\right)e^{\frac{2d'\left(\kappa'+1\right)}{\kappa'+2}}-2e^{\frac{4d'}{(\kappa'+2)}}\right)}$$

For $\log\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)$

$$\begin{aligned}
\log\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right) &\approx \log\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d^{'}}{\kappa^{'}+2}}\right)\\
&+\frac{\partial\log\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)}{\partial d}\left(d-d^{'}\right)+\frac{\partial\log\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)}{\partial \kappa}\left(\kappa-\kappa^{'}\right)\\
&+\frac{\partial^2\log\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)}{\partial d\partial\kappa}\left(d-d^{'}\right)\left(\kappa-\kappa^{'}\right)\\
&+\frac{1}{2}\left(d-d^{'}\right)\frac{\partial^2\log\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)}{\partial d^2}\left(d-d^{'}\right)\\
&+\frac{1}{2}\left(\kappa-\kappa^{'}\right)\frac{\partial^2\log\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)}{\partial \kappa^2}\left(\kappa-\kappa^{'}\right)
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial\log\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)}{\partial d}|_{d=d^{'},\kappa=\kappa^{'}}\\
&=\frac{1}{\frac{1}{16}-\frac{1}{16}e^{-\frac{4d^{'}}{\kappa^{'}+2}}}\frac{\partial\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)}{\partial d}|_{d=d^{'},\kappa=\kappa^{'}}\\
&=\frac{1}{\frac{1}{16}-\frac{1}{16}e^{-\frac{4d^{'}}{\kappa^{'}+2}}}\left(\frac{e^{-\frac{4d^{'}}{\kappa^{'}+2}}}{4\left(\kappa^{'}+2\right)}\right)=\frac{4}{\left(\kappa^{'}+2\right)\left(e^{\frac{4d^{'}}{\left(\kappa^{'}+2\right)}}-1\right)}
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial\log\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)}{\partial \kappa}|_{d=d^{'},\kappa=\kappa^{'}}\\
&=\frac{1}{\frac{1}{16}-\frac{1}{16}e^{-\frac{4d^{'}}{\kappa^{'}+2}}}\frac{\partial\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)}{\partial \kappa}|_{d=d^{'},\kappa=\kappa^{'}}\\
&=-\frac{1}{\frac{1}{16}-\frac{1}{16}e^{-\frac{4d^{'}}{\kappa^{'}+2}}}\left(\frac{d^{'}e^{-\frac{4d^{'}}{\kappa^{'}+2}}}{4\left(\kappa^{'}+2\right)^2}\right)=-\frac{4d^{'}}{\left(\kappa^{'}+2\right)^2\left(e^{\frac{4d^{'}}{\left(\kappa^{'}+2\right)}}-1\right)}
\end{aligned}$$

$$\frac{\partial^2 \log\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)}{\partial d^2}$$

$$= \frac{\partial}{\partial d}\left(\frac{4}{(\kappa'+2)\left(e^{\frac{4d}{(\kappa'+2)}}-1\right)}\right)|_{d=d',\kappa=\kappa'} = -\frac{16e^{\frac{4d'}{(\kappa'+2)}}}{(\kappa'+2)^2\left(e^{\frac{4d'}{(\kappa'+2)}}-1\right)^2}$$

$$\frac{\partial^2 \log\left(\frac{1}{16}-\frac{1}{16}e^{-\frac{4d}{\kappa+2}}\right)}{\partial \kappa^2}|_{d=d',\kappa=\kappa'}$$

$$= \frac{\partial}{\partial \kappa}\left(-\frac{4d'}{(\kappa+2)^2\left(e^{\frac{4d'}{(\kappa+2)}}-1\right)}\right)|_{d=d',\kappa=\kappa'}$$

$$= 4d'\frac{1}{\left((\kappa'+2)^2\left(e^{\frac{4d'}{(\kappa'+2)}}-1\right)\right)^2}\frac{\partial}{\partial \kappa}\left((\kappa+2)^2\left(e^{\frac{4d'}{(\kappa+2)}}-1\right)\right)|_{d=d',\kappa=\kappa'}$$

$$= 4d'\frac{1}{\left((\kappa'+2)^2\left(e^{\frac{4d'}{(\kappa'+2)}}-1\right)\right)^2}\left(2\left(\kappa'+2\right)\left(e^{\frac{4d'}{(\kappa'+2)}}-1\right)-4d'e^{\frac{4d'}{(\kappa'+2)}}\right)$$

$$= \frac{8d'\left(\left(\kappa'-2d'+2\right)e^{\frac{4d'}{(\kappa'+2)}}-\kappa'-2\right)}{(\kappa'+2)^4\left(e^{\frac{4d'}{(\kappa'+2)}}-1\right)^2}$$

$$\frac{\partial^2 \log\left(\frac{1}{16}-\frac{1}{16}e^{-\dfrac{4d}{\kappa+2}}\right)}{\partial d\partial\kappa}|_{d=d',\kappa=\kappa'}$$

$$=\frac{\partial}{\partial\kappa}\left(\frac{4}{(\kappa+2)\left(e^{\dfrac{4d'}{(\kappa+2)}}-1\right)}\right)|_{d=d',\kappa=\kappa'}$$

$$=-\frac{4}{\left((\kappa'+2)\left(e^{\dfrac{4d'}{(\kappa'+2)}}-1\right)\right)^2}\frac{\partial}{\partial\kappa}\left((\kappa+2)\left(e^{\dfrac{4d'}{(\kappa+2)}}-1\right)\right)|_{d=d',\kappa=\kappa'}$$

$$=-\frac{4}{\left((\kappa'+2)\left(e^{\dfrac{4d'}{(\kappa'+2)}}-1\right)\right)^2}\left(-\frac{4d'e^{\dfrac{4d'}{(\kappa'+2)}}}{\kappa'+2}+e^{\dfrac{4d'}{(\kappa'+2)}}-1\right)$$

$$=-\frac{4\left(\left(\kappa'-4d'+2\right)e^{\dfrac{4d'}{(\kappa'+2)}}-\kappa'-2\right)}{(\kappa'+2)^3\left(e^{\dfrac{4d'}{(\kappa'+2)}}-1\right)^2}$$

$$\frac{\partial^2 \log\left(\frac{1}{16}-\frac{1}{16}e^{-\dfrac{4d}{\kappa+2}}\right)}{\partial d\partial\kappa}|_{d=d',\kappa=\kappa'}$$

$$=\frac{\partial}{\partial d}\left(-\frac{4d}{(\kappa'+2)^2\left(e^{\dfrac{4d}{(\kappa'+2)}}-1\right)}\right)=-\frac{4}{(\kappa'+2)^2}\frac{\partial}{\partial d}\left(\frac{d}{e^{\dfrac{4d}{(\kappa'+2)}}-1}\right)|_{d=d',\kappa=\kappa'}$$

$$=\frac{4}{(\kappa'+2)^2\left(e^{\dfrac{4d}{(\kappa'+2)}}-1\right)^2}\left(\frac{4d'e^{\dfrac{4d'}{(\kappa'+2)}}}{\kappa'+2}-e^{\dfrac{4d'}{(\kappa'+2)}}+1\right)$$

$$=\frac{4\left(\left(4d'-\kappa'-2\right)e^{\dfrac{4d'}{(\kappa'+2)}}+\kappa'+2\right)}{(\kappa'+2)^3\left(e^{\dfrac{4d'}{(\kappa'+2)}}-1\right)^2}$$

C. Variational Inference for substitution mapping algorithm

1 Compute the variational function parameters for the number of substitution events

During a time interval, the number of substitution events is distributed according to a Poisson distribution of rate $r_i l_j$ (Lartillot 2006). This is a prior distribution for n_{ij}:

$$p\left(n|r,l\right) = \prod_i \prod_j e^{(-r_i l_j)} \frac{(r_i l_j)^{n_{ij}}}{n_{ij}!} \tag{11}$$

The posterior distribution of n_{ij} during a time interval, given that the initial and final states ($\sigma_i = a$ and $\sigma_f = b$) at the process at time evolution $r_i l_j$, is as follow:

$$p\left(n|a,b,\pi,r,l\right) = \frac{p\left(b|a,\pi,n\right) p\left(n|r,l\right)}{\sum\limits_{n\geq 0} p\left(b|a,\pi,n\right) p\left(n|r,l\right)} \tag{12}$$

We consider two cases which are proposed (Lartillot 2006):

- If $a \neq b$ a truncated Poisson distribution is considered for n_{ij}
- If a = b, the posterior distribution of n_{ij} is computed as follow:

$$\begin{array}{l} p_{a\to b}\left(r_i l_j\right) = p\left(a,b|\pi,r,l\right) = e^{-r_i l_j} + \left(1-e^{-r_i l_j}\right)\pi_b \\ p\left(n>0|a,b,\pi,r,l\right) = \prod\limits_i \prod\limits_j \frac{\pi_b}{p(a,b|\pi,r_i,l_j)} \frac{e^{-r_i l_j}(r_i l_j)^{n_{ij}}}{n_{ij}!} \\ p\left(n=0|a,b,\pi,r,l\right) = \prod\limits_i \prod\limits_j \frac{e^{-r_i l_j}}{p(a,b|\pi,r_i,l_j)} \end{array} \tag{13}$$

The log of the optimized factor to the posterior distribution of having n substitutions along a branch of length l_j is

$$\log q^*(n) = \mathrm{E}_{q(\pi),q(z),q(r),q(l)}\left[\log p\left(a,b,\pi,z,r,l,n\right)\right] + const \tag{14}$$

$$\begin{array}{l} \log q^*(n) = \mathrm{E}_{q(z)}(p(z))\mathrm{E}_{q(\pi)}\left[\log p(\pi)\right] + \mathrm{E}_{q(r),q(l)}\left[\log p\left(n|r,l\right)\right] \\ +\mathrm{E}_{q(\pi),q(r),q(l)}\left[\log\left(p\left(a,b|\pi,r,l\right)\right)\right]\mathrm{E}_{q(z)}(p(z)) + const \end{array} \tag{15}$$

The variational parameters of n are updated by computing variational expectation in equation [15] as follow:

$$\begin{array}{l} \omega_{ij} = \sum\limits_{k=1}^{K_{max}} \sum\limits_{a=1}^{20} \phi_i^k \left[\psi\left(\lambda_a^k\right) - \psi\left(\sum\limits_{a'=1}^{20} \lambda_{a'=1}^k\right)\right] \\ + \left[\psi\left(\zeta_i\right) - \log\left(\zeta_i^{'}\right)\right] + \left[\psi\left(\gamma_j\right) - \log\left(\gamma_j^{'}\right)\right] - \left[\frac{\gamma_j}{\gamma_j^{'}}\frac{\zeta_i}{\zeta_i^{'}}\right] \\ + \sum\limits_{k=1}^{K_{max}} \phi_i^k \mathrm{E}_{q(\pi),q(r),q(l)}\left[\log\left(p\left(a,b|\pi,r,l\right)\right)\right] \end{array} \tag{16}$$

2 Compute the variational function parameters for the number of amino acid types in each category

Similarity, the optimized solution to the posterior distribution for the number of amino acid types, conditional on the states at the ends is computed as follows:

$$\begin{array}{l} \log q^*\left(w_a^k\right) = \sum\limits_i \mathrm{E}_{q(z)}\left[p(z_i^k)\right]\mathrm{E}_{q(\pi)}\left[\log\left(\pi_a^k\right)\right] + \sum\limits_i \sum\limits_j \mathrm{E}_{q(r),q_(l)} \log\left[1-e^{-r_i l_j}\right] \\ + \sum\limits_i \sum\limits_j \mathrm{E}_{q(z)}\left[p(z_i^k)\right]\mathrm{E}_{q(\pi),q(r),q(l)}\left[\log\left(p\left(a,b|\pi,r_i,l_j\right)\right)\right] \end{array} \tag{17}$$

The variational parameters of ι_a^k are updated as follows:

$$\begin{array}{l} \iota_a^k = \sum\limits_i \phi_i^k \left[\psi\left(\lambda_a^k\right) - \psi\left(\sum\limits_{a'=1}^{20} \lambda_{a'}^k\right)\right] + \sum\limits_i \sum\limits_j \mathrm{E}_{q(r),q_(l)} \log\left[1-e^{-r_i l_j}\right] \\ + \sum\limits_i \sum\limits_j \phi_i^k \mathrm{E}_{q(\pi),q(r),q(l)}\left[\log\left(p\left(a,b|\pi,r_i,l_j\right)\right)\right] \end{array} \tag{18}$$

The expected logarithm of the functions $\mathrm{E}_{q(\pi),q(r),q(l)}\left[\log\left(p\left(a,b|\pi,r_i,l_j\right)\right)\right]$ and $\mathrm{E}_{q(r),q(l)}\left[\log\left(1-e^{-r_il_j}\right)\right]$ in equation (16)(18) have not the closed form. Thus, calculation of these equations are analytically intractable. To use standard form of variational inference (Bishop. 2006) or stochastic variational inference (Hoffman et al. 2013), we need a closed-form expression. In order to overcome these problems, we consider two approaches:

- By applying a first-order Taylor expansion to preserve a bound, intractable expectations are avoided. We explain specifically this approaches to approximate expectations as follow.

The product of r_il_j is denoted t_{ij}, firstly, we consider a first-order Taylor expansion of $\log\left(p\left(a,b|\pi,t_{ij}\right)\right)$ for π_a and t_{ij} at $\pi_a^{'}=\dfrac{\upsilon_a}{\sum\limits_{a^{'}=1}^{20}\upsilon_{a^{'}}}$ and $t_{ij}^{'}=\dfrac{1}{\beta}\dfrac{\alpha}{\alpha}=\dfrac{1}{\beta}$

$$\begin{aligned}\log\left[e^{-t_{ij}}+\left(1-e^{-t_{ij}}\right)\pi_a\right]&\approx\log\left[e^{-t_{ij}^{'}}+\left(1-e^{-t_{ij}^{'}}\right)\pi_a^{'}\right]\\&+\frac{\partial\log\left[e^{-t_{ij}}+\left(1-e^{-t_{ij}}\right)\pi_a^{'}\right]}{\partial t_{ij}}\left(t_{ij}-t_{ij}^{'}\right)\\&+\frac{\partial\log\left[e^{-t_{ij}^{'}}+\left(1-e^{-t_{ij}^{'}}\right)\pi_a\right]}{\partial\pi_a}\left(\pi_a-\pi_a^{'}\right)\end{aligned}\tag{19}$$

$$\begin{aligned}&\frac{\partial\log\left[e^{-t_{ij}}+\left(1-e^{-t_{ij}}\right)\pi_a^{'}\right]}{\partial t_{ij}}\Big|_{t_{ij}=t_{ij}^{'}}=\frac{1}{e^{-t_{ij}^{'}}+\left(1-e^{-t_{ij}^{'}}\right)\pi_a^{'}}\frac{\partial\left[e^{-t_{ij}}+\left(1-e^{-t_{ij}}\right)\pi_a^{'}\right]}{\partial t_{ij}}\\&=\frac{\pi_a^{'}e^{-t_{ij}^{'}}-e^{-t_{ij}^{'}}}{e^{-t_{ij}^{'}}+\left(1-e^{-t_{ij}^{'}}\right)\pi_a^{'}}=\frac{\pi_a^{'}-1}{\pi_a^{'}e^{t_{ij}^{'}}-\pi_a^{'}+1}\end{aligned}\tag{20}$$

$$\begin{aligned}&\frac{\partial\log\left[e^{-t_{ij}^{'}}+\left(1-e^{-t_{ij}^{'}}\right)\pi_a\right]}{\partial\pi_a}\Big|_{\pi=\pi^{'}}=\frac{1}{e^{-t_{ij}^{'}}+\left(1-e^{-t_{ij}^{'}}\right)\pi_a^{'}}\frac{\partial\left[e^{-t_{ij}^{'}}+\left(1-e^{-t_{ij}^{'}}\right)\pi\right]}{\partial\pi}\\&=\frac{1-e^{-t_{ij}^{'}}}{e^{-t_{ij}^{'}}+\left(1-e^{-t_{ij}^{'}}\right)\pi_a^{'}}=\frac{e^{t_{ij}^{'}}-1}{\left(e^{t_{ij}^{'}}-1\right)\pi_a^{'}+1}\end{aligned}\tag{21}$$

Substitute equation (20)(21) for equation (19), we have the expected value as follow:

$$\begin{aligned}&\mathrm{E}_{q(r_i),q(l_j),q(\pi)}\left(\log\left[e^{-r_il_j}+\left(1-e^{-r_il_j}\right)\pi\right]\right)\\&\approx\sum_a\log\left[e^{-\frac{1}{\beta}}+\left(1-e^{-\frac{1}{\beta}}\right)\pi_a^{'}\right]\\&+\sum_a\frac{\pi_a^{'}-1}{\pi_a^{'}e^{\frac{1}{\beta}}-\pi_a^{'}+1}\left(\mathrm{E}_{q(r_i)}\left(r_i\right)\mathrm{E}_{q(l_j)}\left(l_j\right)-\frac{1}{\beta}\right)\\&+\sum_a\frac{e^{\frac{1}{\beta}}-1}{\left(e^{\frac{1}{\beta}}-1\right)\pi_a^{'}+1}\left(\mathrm{E}_{q(\pi)}(\pi)-\pi_a^{'}\right)\\&\approx\sum_a\log\left[e^{-\frac{1}{\beta}}+\left(1-e^{-\frac{1}{\beta}}\right)\pi_a^{'}\right]+\sum_a\frac{\pi_a^{'}-1}{\pi_a^{'}e^{\frac{1}{\beta}}-\pi_a^{'}+1}\left(\frac{\zeta_i}{\zeta_i^{'}}\frac{\gamma_j}{\gamma_j^{'}}-\frac{1}{\beta}\right)\\&+\sum_{k=1}^{K_{max}}\sum_{a=1}^{20}\frac{e^{\frac{1}{\beta}}-1}{\left(e^{\frac{1}{\beta}}-1\right)\pi_a^{'}+1}\left(\left[\psi\left(\lambda_a^k\right)-\psi\left(\sum_{a^{'}=1}^{20}\lambda_{a^{'}=1}^k\right)\right]-\pi_a^{'}\right)\end{aligned}\tag{22}$$

We consider first-order Taylor expansion for $\log\left(1-e^{-t_{ij}}\right)$ for t_{ij} at $t_{ij}^{'}=\dfrac{1}{\beta}$

$$\log\left(1-e^{-t_{ij}}\right) \approx \log\left(1-e^{-t_{ij}^{'}}\right) + \frac{\partial \log\left(1-e^{-t_{ij}}\right)}{\partial t_{ij}}\left(t_{ij}-t_{ij}^{'}\right) \tag{23}$$

$$\frac{\partial \log\left(1-e^{-t_{ij}}\right)}{\partial t_{ij}}|_{t_{ij}=t_{ij}^{'}} = \frac{1}{1-e^{-t_{ij}^{'}}}\frac{\partial\left(1-e^{-t_{ij}}\right)}{\partial t_{ij}} = \frac{e^{-t_{ij}^{'}}}{1-e^{-t_{ij}^{'}}} = \frac{1}{e^{t_{ij}^{'}}-1} \tag{24}$$

Substitute the result of equation (24) to equation (23), we obtain the value value as follow:

$$\begin{aligned}
&\mathrm{E}_{q(r_i)q(l_j)}\left[\log\left(1-e^{-r_i l_j}\right)\right] \\
&\approx \log\left(1-e^{-\frac{1}{\beta}}\right) + \frac{1}{e^{\frac{1}{\beta}}-1}\left(\mathrm{E}_{q(r_i)}\left(r_i\right)\mathrm{E}_{q(l_j)}\left(l_j\right)-\frac{1}{\beta}\right) \\
&\approx \log\left(1-e^{-\frac{1}{\beta}}\right) + \frac{1}{e^{\frac{1}{\beta}}-1}\left(\frac{\zeta_i}{\zeta_i^{'}}\frac{\gamma_j}{\gamma_j^{'}}-\frac{1}{\beta}\right)
\end{aligned} \tag{25}$$

By using the results of equation (22)(25), the variational parameters of the number of substitution and the number of type of amino acid are updated as follow:

$$\begin{aligned}
&\omega_{ij} \approx \sum_{k=1}^{K_{max}}\sum_{a=1}^{20}\phi_i^k\left[\psi\left(\lambda_a^k\right)-\psi\left(\sum_{a'=1}^{20}\lambda_{a'=1}^k\right)\right] \\
&+\left[\psi\left(\zeta_i\right)-\log\left(\zeta_i^{'}\right)\right]+\left[\psi\left(\gamma_j\right)-\log\left(\gamma_j^{'}\right)\right]-\left[\frac{\gamma_j}{\gamma_j^{'}}\frac{\zeta_i}{\zeta_i^{'}}\right] \\
&+\sum_a\log\left[e^{-\frac{1}{\beta}}+\left(1-e^{-\frac{1}{\beta}}\right)\pi_a^{'}\right]+\sum_a\frac{\pi_a^{'}-1}{\pi_a^{'}e^{\frac{1}{\beta}}-\pi_a^{'}+1}\left(\frac{\zeta_i}{\zeta_i^{'}}\frac{\gamma_j}{\gamma_j^{'}}-\frac{1}{\beta}\right) \\
&+\sum_{k=1}^{K_{max}}\sum_{a=1}^{20}\phi_i^k\left(\frac{e^{\frac{1}{\beta}}-1}{\left(e^{\frac{1}{\beta}}-1\right)\pi_a^{'}+1}\right)\left(\left[\psi\left(\lambda_a^k\right)-\psi\left(\sum_{a'=1}^{20}\lambda_{a'=1}^k\right)\right]-\pi_a^{'}\right)
\end{aligned} \tag{26}$$

$$\begin{aligned}
&\iota_a^k \approx \sum_i\phi_i^k\left[\psi\left(\lambda_a^k\right)-\psi\left(\sum_{a'=1}^{20}\lambda_{a'}^k\right)\right] \\
&+\log\left(1-e^{-\frac{1}{\beta}}\right)+\left(\frac{1}{e^{\frac{1}{\beta}}-1}\right)\sum_i\sum_j\left(\frac{\zeta_i}{\zeta_i^{'}}\frac{\gamma_j}{\gamma_j^{'}}-\frac{1}{\beta}\right) \\
&+\sum_a\log\left[e^{-\frac{1}{\beta}}+\left(1-e^{-\frac{1}{\beta}}\right)\pi_a^{'}\right]+\sum_a\frac{\pi_a^{'}-1}{\pi_a^{'}e^{\frac{1}{\beta}}-\pi_a^{'}+1}\left(\frac{\zeta_i}{\zeta_i^{'}}\frac{\gamma_j}{\gamma_j^{'}}-\frac{1}{\beta}\right) \\
&+\sum_i\phi_i^k\sum_{a=1}^{20}\left(\frac{e^{\frac{1}{\beta}}-1}{\left(e^{\frac{1}{\beta}}-1\right)\pi_a^{'}+1}\right)\left(\left[\psi\left(\lambda_a^k\right)-\psi\left(\sum_{a'=1}^{20}\lambda_{a'=1}^k\right)\right]-\pi_a^{'}\right)
\end{aligned} \tag{27}$$

3 Reference

1. Bishop. C. M. 2006. Pattern Recognition and Machine Learning. Springer, 1 edition.
2. Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368–376.

3. Blei, D and Jordan, M. 2006. Variational inference for Dirichlet process mixtures. Journal of Bayesian Analysis, 1:121–144.

4. Gopalan, P., Hao, W., Blei, D., Storey, J. 2016. Scaling probabilistic models of genetic variation to millions of humans. Nat Genet. 48, 1587-1590.

5. Hoffman, M., Blei, D., Wang, C. & Paisley, J. 2013. Stochastic variational inference. J. Mach. Learn. Res. 14, 1303–1347.

6. Jordan, M., Ghahramani, Z., Jaakkola, T. & Saul, L. 1999. Introduction to variational methods for graphical models. Mach. Learn. 37, 183–233.

7. Lartillot N., Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21:1095–1109.

8. Lartillot N. 2006. Conjugate Gibbs sampling for Bayesian phylogenetic models. J. Comput. Biol. 13:1701–1722.

9. Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst. Biol.. 62, 611–615.

10. Nielsen R. 2002. Mapping mutations on phylogenies. Syst. Biol. 51: 729–739.

11. Quang L.S., Gascuel O., Lartillot N. 2008. Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics 24: 2317–2323.

12. Robbins, H. and Monro, S. 1951. A stochastic approximation method. The Annals of Mathematical Statistics 22, 400–407.

13. Wainwright, M. & Jordan, M. 2008. Graphical models, exponential families, and variational inference. Found. Trends Mach. Learn. 1, 1–305.