Modeling mammalian methylome evolution

Jianghan Qu [1] and Andrew D. Smith[1]

[1]Molecular and Computational Biology Section, Division of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA

1 State space and units of measurement for DNA methylation

DNA methylation is often discussed in terms of levels at individual CpG sites, and the level reflects the fraction of cells that have the discrete methyl mark at that site (more accurately molecules with the mark, in the case of non-haploid cells). In multiple studies since 2009, when considering cells that are relatively pure in terms of phenotype, the methylation levels have been observed to fall into two categories: high and low levels. This is seen in the global bi-modal distributions of methylation levels, and when one observes profiles of DNA methylation in a genome browser. There are special cases of intermediate methylation levels, for example cancers have large domains of partial methylation. In addition, imprinted loci have intermediate methylation (one allele methylated through an imprinting control region in most somatic cells). However, for the vast majority of the sites, major phenotypic differences among healthy cells usually involve methylation changing from low to high, or from high to low. For this reason, all our modeling is in terms of low and high methylation states, and we use the corresponding state space $\{0, 1\}$. This allows for a distribution of the observed levels associated with the "low" methylation state, and another distribution for the observed levels at sites occupying the "high" state.

Although this restriction is justified, our modeling approach still allows for an interpretation that is consistent with the fact that methylation is usually measured as levels between 0 and 1. One may consider that level as (an unbiased estimator for) the probability that a randomly sampled molecule would have a methyl group at the site. In our modeling we ultimately make use of probabilities over the state space, which

behave very much like a continuous level. At the same time, the discrete state space is highly convenient, and the preponderance of evidence indicates that in most cases the state carries almost the same information as the level (and may be more robust to artifacts and random noise).

Current technology for measuring DNA methylation can interrogate single nucleotides (in mammals, with specific exceptions, these are cytosines of CpG dinucleotides). In modeling how methylation status relates between species we must use measurements that can be considered orthologous. Part of our analyses uses 200 base bins, assigning a methylation state to the bin, and using the average of individual site methylation levels to obtain a level for the bin, when necessary. We also devise a strategy to model individual CpG sites directly, even when those sites are not mutually orthologous within all considered species. Below we explain our modeling approach in terms of CpG sites, but the same approaches can be applied for any kind of unit to which a methylation level may be assigned.

2 Model identifiability

Markov models on evolutionary trees have been proved useful and widely applied in phylogenetic studies. An evolutionary Markov model consists of a tree topology and a set of transition probability matrices on the branches of the tree. Identifiability of the tree topology from the distribution of character states at terminal nodes have been established (Chang & Hartigan 1992, Steel et al. 1995). Under some mild conditions, the full model is identifiable from knowledge of the joint distribution of character states at triples of terminal nodes of the tree (Chang 1996). It is a common practice to model the evolution process with a time-reversible Markov process, which makes the phylogenetic model satisfy conditions for full model identifiability (Liò & Goldman 1998). A recent application of phylogenetic model to DNA methylation changes in the context of cell lineage differentiation also assumed reversibility (Capra & Kostka 2014). However, preliminary results from our comparative analysis of sperm methylomes across species have shown that the methylome evolution might be directional and have not reached equilibrium during recent mammalian evolution. Therefore, we did not require the Markov process modeling the epigenome evolution to be reversible. Instead, in order to satisfy model identifiability conditions, we assumed that one of the two branches at the root is effectively 0. Based on previous phylogenetic and archaeological dating of species divergence time, we set the shorter branch, between the last common ancestor (LCA) of human and dog and the LCA of human and mouse,

to 0 to have minimal deviation from the theoretical binary tree. This is equivalent to assuming a rooted semi-binary tree, where the root has 3 children and other internal nodes have 2 children.

3 Phylo-epigenetic model with independent sites

We begin by assuming that methylation levels at CpG sites are independent and identically distributed, following a mixture of two beta distributions. The two components of the mixture distribution correspond to the low and high methylation states. Given the state of an individual site, the WGBS read proportion follows a beta-binomial distribution. We obtain the maximum likelihood estimates (MLE) of the mixing proportion and beta distribution parameters using an Expectation Maximization (EM) algorithm. The M-step of the EM algorithm is as previously described in a hidden Markov model setting for identifying hypomethylated regions (Molaro et al. 2011). Then we calculate posterior probabilities of individual CpG sites being in the "low" state. These hypomethylation probabilities serve as input data for the phyloepigenetic model assuming site-independence.

Our goal is to model the evolution of DNA methylation states in multiple extant species with a common ancestor. We first focus on a single-site to derive the likelihood function, and then extend the model to multiple sites but for which epigenomic evolution is assumed to be independent.

We assume that the methylation state at a single CpG site evolves according to a two-state continuous-time Markov process. Let $\pi = (\pi_0, \pi_1)$ be the initial distribution of the methylation state at the root node, and let the transition rate matrix be

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \eta & -\eta \end{bmatrix}.$$

The transition probability matrix between two time points separated by time t is $P(t) = \exp(Qt)$, which is determined by two terms $(t(\lambda + \eta), \lambda/\eta)$. Let $\lambda + \eta = 1$, so that the mutation rate and branch length parameters are identifiable.

Let $\tau = \{V, E\}$ be the phylogenetic tree with known topology and unknown branch lengths, where V is the set of known vertices and E is the set of branches with unknown lengths. The model parameter space is thus

$$\Theta = \{E, \lambda, \pi\}.$$

We require some notation for representing nodes and their relationships in the tree. We use r to denote the root node of the phylogenetic tree, and use u, v and w to denote 3 consecutive nodes on a lineage, such that $(u, v) \in E$ and $(v, w) \in E$. Let random variable $s(u) \in \{0, 1\}$ be the state of node u. Let ℓ_v be the length of edge $(u, v) \in E$. Let $X(u)$ be the set of methylation states at leaf nodes that are descendents of u. Thus, the observed data is represented by $X(r)$. We use j and k to denote methylation states. In general, we will use j to denote the state at the parent of a node whose state is denoted by k.

For node v, whose parent u has methylation state $s(u) = j$, the probability of observing states $X(v)$ at terminal descendents of node v is

$$p_j(v) = \Pr(X(v)|s(u) = j, \Theta).$$

For notational convenience we define

$$q_k(v) = \begin{cases} \Pr(s(v) = k) & \text{if } v \text{ is a leaf node,} \\ \prod_{w \in \text{child}(v)} p_k(w) & \text{otherwise,} \end{cases}$$

where $\Pr(s(v) = k)$ is 0 or 1 when methylation state is observed, and between $(0, 1)$ when the observed data is a continuous level representing a probability distribution of the state space.

We can then write the probability $p_j(v)$ as the recurrence

$$p_j(v) = \sum_k \Big[P(\ell_v)_{jk} \times q_k(v) \Big]. \tag{1}$$

The likelihood of the observed data for a single CpG site is thus

$$L(\Theta; X) = \Pr(X(r)|\Theta) = \sum_{j \in \{0,1\}} \pi_j q_j(r). \tag{2}$$

The recurrence in (1) is the basis of Felsenstein's pruning algorithm for efficiently computing the likelihood of a tree topology, branch lengths and transition rate, given data at leaf nodes (Felsenstein 1981).

Moving from single site to multiple sites, let N be the total number of sites in the methylome. Let $X_n (1 \leq n \leq N)$ be the set of methylation states associated with leaf nodes at site n. $X = X_1 \dots X_N$

denote the observed methylation states at all leaf nodes. Since methylation states at distinct sites of the methylome evolve independently, the likelihood for observed data at multiple sites is

$$L(\Theta|X) = \prod_{n=1}^{N} L(\Theta|X_n).$$

In our implementation, we optimize parameters using gradient descent. The likelihood and gradients are recursively computed in the same spirit of the pruning algorithm (Felsenstein 1981). With the MLE of model parameters, we then compute the maximum likelihood reconstruction of the HME at each CpG site. The reconstruction can be achieved through a dynamic programming algorithm with time complexity linear to the number of nodes in the tree (Pupko et al. 2000).

4 Phylo-epigenetic model with dependent sites

We develop a model to allow for two processes that jointly describe the observed mammalian methylome: one is methylation state inheritance from ancestral species, and the other is the correlation observed between neighboring sites within a species.

The inheritance process is defined by Q, or equivalently $P(t)$, as previously introduced. The correlation between neighboring CpG sites in a species is described with a discrete-time (corresponding to CpG sites) Markov chain over the state space $\{0, 1\}$. Let the initial distribution be π in the root species. The transition probability matrix,

$$G = \begin{bmatrix} g_0 & 1-g_0 \\ 1-g_1 & g_1 \end{bmatrix}$$

is assumed to be homogeneous in all species (this assumption can be relaxed).

As in Section 3, we use $\tau = \{V, E\}$ to denote the phylogenetic tree relating species, with known topology and unknown branch lengths. The model parameter space is thus

$$\Theta = \{\tau, Q, G, \pi_0\}.$$

The two Markov processes are combined in the following way to model the evolution of N consecutive

CpG sites in a genomic region. The CpG sites are ordered from 1 to N by their occurrences from the 5' end of the '+' strand of a chosen primary reference genome (human reference genome in this study).

Consider two neighboring CpG sites $n-1$ and n, and a pair of nodes (u, v) linked by a branch in the phylogenetic tree, where u is the ancestor of v. Let random variable v_n be the methylation state of site n in species v. The conditional distribution of v_j given the methylation state of previous CpG site v_{n-1} and the ancestral methylation state u_n is defined as:

$$p_v(i,j,k) = \Pr(v_n = k | v_{n-1} = i, u_n = j) = \frac{G_{ik} P(\ell_v)_{jk}}{\sum_{k'=0,1} G_{ik} P(\ell_v)_{jk'}},\ \forall i,j,k \in \{0,1\}. \tag{3}$$

Special cases involve sites at position $n = 1$, and sites in the root node. The root methylome is modeled only with the discrete-time Markov chain for site-interdependence, $M_{dep}\{\pi, G\}$. For sites at position $n = 0$, their methylation state evolution is modeled only with the continuous-time Markov process for inheritance and mutation, $M_{inh}\{\pi, Q, \tau\}$.

Complete data likelihood We examine the the complete data likelihood, from which we later derive approximation method for model learning. Assume the methylation states at all sites in each external and internal species are observed, denoted by O. Given model parameters $\Theta = \{\tau, Q, G, \pi_0\}$, the complete data likelihood is

$$L(\Theta; O) = \Pr(O|\Theta) = \Pr(O_1|\Theta) \prod_{n=2}^{N} G_{o_{n-1}(r) o_n(r)} \prod_{u \in I} \prod_{v \in \text{child}(u)} p_v(O_{n-1}(v), O_n(u), O_n(v))$$

where *I* is the set of internal nodes, and $O_n(v)$ is the methylation state of node v at site n.

$$\begin{aligned} \log L(\Theta; O) = & \log \pi_{O_1(r)} + \sum_{v \neq r} \sum_{j,k \in \{0,1\}} w_{jk}(v) P(\ell_v)_{jk} + \\ & \sum_{i,k \in \{0,1\}} w_{ik} \log(G_{ij}) + \sum_{v \neq r} \sum_{i,j,k \in \{0,1\}} w_{ijk}(v) p_v(i,j,k) \end{aligned} \tag{4}$$

where

$$
\begin{aligned}
w_{ijk}(v) &= \sum_{n=2}^{N} I\{O_{n-1}(v) = i, O_n(u) = j, O_n(v) = k\} \\
w_{jk}(v) &= I\{O_n(u) = j, O_n(v) = k\} \\
w_{ik} &= \sum_{n=2}^{N} I\{O_{n-1}(r) = i, O_n(r) = k\}
\end{aligned} \tag{5}
$$

where I is an indicator function. Therefore, $W = \{w_{jk}(v), w_{ijk}(v), w_{ik} : i, j, k \in \{0, 1\}, v \in V \backslash \{r\}\}$ are the sufficient statistics for the model parameters. The MLE for model parameters can be efficiently obtained through numerical methods, such as gradient descent, given complete data.

Hidden Markov model representation The ancestral methylomes are unobserved, which we hope to infer using observed data from the extant species and their phylogenetic relationship. We can interpret our interdependent-site phyloepigenetic model as a hidden Markov model over the state space

$$H = \{h \in \{0, 1\}^{|V|} : \text{all possible combination of methylation states over the phylogenetic tree}\}.$$

Let the H_n be a random variable denoting the HME at site n. Each state specifies the history of the methylation evolution a single genomic site, and thus we call it "History of Methylation Evolution" (HME). The transition probability between two HMEs can be calculated using quantities defined in the previous section. Let $h, h' \in H$ be the HMEs at two neighboring sites in the genome. Let $u(h)$ be the methylation state at node u as specified by HME h. Let v be an internal or leaf node in the phylogenetic tree, and let its parent node be u. The transition probability from HME h to h' is

$$\Pr(H_{n+1} = h' | H_n = h) = G_{r(h)r(h')} \prod_{(u,v) \in E} p_v(v(h), u(h'), v(h')), \tag{6}$$

where $p_v(v(h), u(h'), v(h'))$ is defined in (3). Use $A = \{a_{hh'} : h, h' \in H\}$ to denote the transition probability matrix between HME states.

The data likelihood calculation can be written as

$$L(\Theta; X) = \sum_{H_1,\ldots,H_N \in H} \Pr(H_1) \Pr(X_1|H_1) \prod_{n=1}^{N-1} \Pr(H_{n+1}|H_n) \Pr(X_{n+1}|H_{n+1}). \tag{7}$$

Bayesian network interpretation The complete data likelihood in (4) is factorized by conditional probabilities, as our model naturally defines a dynamic Bayesian network (DBN). The history of evolution (HME) of each site in the methylome corresponds to a time-slice in DBN. Each site in each species corresponds to a node, a random variable for methylation state, in the graph. The interdependence between neighboring sites and inheritance relationship between ancestor and descendant sites define directed edges between nodes. With observations made at a subset of the nodes in the network (nodes associated with leaf species), we aim to learn the model parameters. Given the sufficient statistics in (5), the MLE of model parameters can be derived by maximizing (4). Given the model parameters θ, and observations from the leaf nodes X, theoretically we can derive the joint distribution of the hidden variables (internal node methylation states Z) conditional on the observed leaf data. We refer to the probability distributions conditional on the observed data as posterior distributions. Using the joint posterior distribution we can obtain expected values of the sufficient statistics $E_{Z|X,\theta}W$. These two processes are exactly the two steps in Expectation-Maximization (EM) algorithm.

E-step:

$$\begin{aligned} Q(\theta|\theta^{(t)}) =& E_{Z|X,\theta^{(t)}} \log(L(\theta; X, Z)) \\ =& \sum_{v \neq r} \sum_{j,k \in \{0,1\}} E(w_{jk}(v)|X, \theta^{(t)}) P(\ell_v)_{jk} + \\ & \sum_{i,k \in \{0,1\}} E(w_{ik}|X, \theta^{(t)}) \log(G_{ik}) + \\ & \sum_{v \neq r} \sum_{i,j,k \in \{0,1\}} E(w_{ijk}(v)|X, \theta^{(t)}) p_v(i,j,k) \end{aligned} \tag{8}$$

M-step:

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmin}} Q(\theta|\theta^{(t)})$$

where $L(\theta; X, Z)$ is the complete data likelihood defined in (4). M-step can be effectively solved by gradient-descent or other numerical methods.

Metropolis-Hasting algorithm For E-step, we use Markov Chain Monte Carlo (MCMC) method, Metropolis-Hasting algorithm in particular, to obtain the expectation of the sufficient statistics given model parameters and observed data at leaf nodes.

- Start from a specific initiation of the states at all nodes in internal species, denoted with Z_0.
- Iterate through all nodes in the graph, from site 1 to site N, and for each site according to a post-order traversal of species in the phylogenetic tree. At each node z in the graph, we make a proposal to flip its state, i.e. $z^{prop} = 1 - z^{(t)}$. We accept the proposal $z^{(t)} = z^{prop}$ with probability

$$\alpha = min\{1, \frac{P(z^{prop}|MB(z, t-1))}{P(z^{(t-1)}|MB(z, t-1))}\},$$

  where $MB(v)$ denotes the Markov blanket of node v in DBN, and $MB(z, t)$ denotes the value of these variables in the t^{th} sample in MCMC. If the proposal is rejected, let $z^{(t)} = z^{(t-1)}$. After we iterate through all the hidden nodes in the graph, we have generated an new sample Z_t.
- From sample $\{X, Z_t\}$, we can obtain W_t as an estimator of $E_{Z|X,\theta}W$.

Measure chain convergence. The MCMC is guaranteed to converge to the target distribution $\Pr(Z|X, \theta)$. To determine convergence time, we let two chains from different starting points run independently and measure the difference between the two chains each time new samples are generated. Our goal is to approximate $E_{Z|X,\theta}W$. Let W_t' and W_t'' be the estimators from the two chains. Notice that

$$\sum_{ijk} w_{ijk}(v) = N, \ \ \forall v \in V \backslash \{r\}),$$

therefore $\frac{1}{N}\{w_{ijk}(v)\}$ is a probability vector, which for simplicity we also denote as $w_{ijk}(v)$. To measure the difference between W_t' and W_t'', we use Kullback-Leibler(KL) divergence between the probability vectors

associated with each internal node v:

$$KL_t(v) = D_{KL}(w'_{ijk}(v)||w''_{ijk}(v)), \;\; v \in V\backslash\{r\}.$$

For a small $\epsilon \in (0, 1)$, we consider the two chains to be mixed at time

$$t_{mix} = \underset{t}{\text{argmin}}\{t : \max\{KL_t(v) : v \in V\backslash\{r\}\} < \epsilon\}.$$

We stop the chains at time t_{mix}, and use $W'_{t_{mix}}$ as an approximation for $E_{Z|X,\theta}W$. In our implementation, $\epsilon = 1e-4$.

Computing approximate posterior probability for individual sites We use the Markov blanket of each node to update its marginal posterior. Let v_n be the methylation state at the n^{th} position in species v in the phylogenetic tree. Let $B(v_n)$ be the set of joint states for nodes in v_n's Markov blanket. The joint distribution of nodes in the Markov blanket is approximated with the product of their marginal distributions. The approximated probability distribution is denoted as $p_b(v_n), b \in B(v_n)$ below.

There are 9 types of nodes in the network, for each of which we describe the composition of the Markov blanket, and the approximation procedure.

Case 1: Root species r, site $n = 1$. Its Markov blanket includes three children $l_n, m_n r_n$, and the neighboring site r_{n+1}. For each state of the Markov blanket, the normalized conditional hypomethylation probability is calculated by

$$\Pr(r_n = 0|b) \propto \prod_{c \in \text{child}(r)} \Pr(c_n|r_n = 0)$$

And the probability distribution of r_n's Markov blanket $B(r_n)$ is

$$p_b(r_n) = \prod_{c \in \text{child}(r)} p_{c_n} \times p_{r_{n+1}}, \forall\, b \in B(v_n).$$

Case 2: Root species r, site $n = N$. The Markov blanket includes $\{c_m : c \in \text{child}(r), m = n-1, n.\}$ and

r_{n-1}.

$$p_b(r_n) = \prod_{c \in \text{child}(r)} p_{c_n} p_{c_{n-1}}$$

$$\Pr(r_n = 0|b) \propto \Pr(r_n = 0|r_{n-1}) \prod_{c \in \text{child}(r)} p_c(c_{n-1}, 0, c_n)$$

Case 3: Root species r, site $1 < n < N$. The Markov blanket includes children states $\{c_n, c_{n-1} : c \in \text{child}(r)\}$, and neighboring states $\{r_{n-1}, r_{n+1}\}$.

$$p_b(r_n) = \prod_{c \in \text{child}(r)} p_{c_n} p_{c_{n-1}} \times p_{r_{n-1}} p_{r_{n+1}}.$$

$$\Pr(r_n = 0|b) \propto \Pr(r_n = 0|r_{n-1}) \Pr(r_{n+1}|r_n = 0) \prod_{c \in \text{child}(r)} p_c(c_{n-1}, 0, c_n)$$

Case 4: Species $v \in \mathcal{L}$ is a leaf species, and site $n = 1$. Its Markov blanket includes v_{n+1}, and u_n, u_{n+1}, where u is the parent of node v.

$$p_b(v_n) = p_{v_{n+1}} p_{u_n} p_{u_{n+1}}$$

$$\Pr(v_n = 0|b) \propto \Pr(v_n = 0|u_n) p_v(0, u_{n+1}, v_{n+1})$$

Case 5: Species $v \in \mathcal{L}$ is a leaf species, and $n = N$. Its Markov blanket includes v_{n-1} and u_n, where u is the parent of node v.

$$p_b(v_n) = p_{v_{n-1}} p_{u_n}$$

$$\Pr(v_n = 0|b) \propto p_v(v_{n-1}, u_n, 0)$$

Case 6: Species $v \in \mathcal{L}$ is a leaf species, and site $1 < n < N$. The Markov blanket includes v_{n-1}, v_{n+1}, u_n

and u_{n+1}.

$$p_b(v_n) = p_{v_{n-1}} p_{v_{n+1}} p_{u_n} p_{u_{n+1}}$$
$$\Pr(v_n = 0|b) \propto p_v(v_{n-1}, u_n, 0) p_v(0, u_{n+1}, v_{n+1})$$

Case 7: Species $v \in V \backslash \{\{r\} \cup \mathcal{L}\}$ is neither root nor leaf, and site $n = 1$. Its Markov blanket includes $\{c_n : c \in \text{child}(v)\}$, v_{n+1}, u_n and u_{n+1}.

$$p_b(v_n) = p_{v_{n+1}} p_{u_n} p_{u_{n+1}} \prod_{c \in \text{child}(v)} p_{c_n}$$
$$\Pr(v_n = 0|b) \propto \Pr(v_n = 0|u_n) p_v(0, u_{n+1}, v_{n+1}) \prod_{c \in \text{child}(v)} \Pr(c_n = 0|v_n = u)$$

Case 8: Species $v \in V \backslash \{\{r\} \cup \mathcal{L}\}$ is neither root nor leaf, and site $n = N$. Its Markov blanket includes $\{c_n, c_{n-1} : c \in \text{child}(v)\}$, v_{n-1} and u_n.

$$p_b(v_n) = p_{v_{n-1}} p_{u_n} \prod_{c \in \text{child}(v)} p_{c_{n-1}} p_{c_n}$$
$$\Pr(v_n = 0|b) \propto p_v(v_{n-1}, u_n, 0) \prod_{c \in \text{child}(v)} p_c(c_{n-1}, 0, c_n)$$

Case 9: Species $v \in V \backslash \{\{r\} \cup \mathcal{L}\}$ is neither root nor leaf, and site $1 < n < N$. The Markov blanket of v_n includes v_{n-1}, $\{c_n, c_{n-1} : c \in \text{child}(v)\}$, v_{n+1}, u_n and u_{n+1}.

$$p_b(v_n) = p_{v_{n-1}} \prod_{c \in \text{child}(v)} p_{c_{n-1}} p_{c_n} p_{v_{n+1}} p_{u_n} p_{u_{n+1}}$$
$$\Pr(v_n = 0|b) \propto p_v(v_{n-1}, u_n, 0) p_v(0, u_{n+1}, v_{n+1}) \prod_{c \in \text{child}(v)} p_c(c_{n-1}, 0, c_n)$$

The posterior hypomethylation probability of site n of species v can be approximated with

$$p'_{v_n} = \sum_{b \in B(v_n)} p_b \Pr(v_n = 0|b).$$

Summary of model inference procedure Together, the procedure for model parameter estimation and ancestral state reconstruction is summarized as below:

1. Choose start point for model parameters $\theta^{(t)}$.

2. Iterate the following EM procedure

   - E-step: Use Metropolis-Hasting algorithm to approximate $E_{Z|X,\theta^{(t)}} W(X, Z)$.
   - M-step: update model parameters to $\theta^{(t+1)}$.

   until convergence: $||\theta^{(t+1)} - \theta^{(t)}|| < \epsilon$.

3. Estimate posterior hypomethylation probability at internal nodes using Markov blanket. Use MAP state as the inference for ancestral methylation states.

References

Capra, J. A. & Kostka, D. (2014), 'Modeling DNA methylation dynamics with approaches from phylogenetics', *Bioinformatics* **30**(17), i408–i414.

Chang, J. T. (1996), 'Full reconstruction of markov models on evolutionary trees: identifiability and consistency', *Mathematical Biosciences* **137**(1), 51–73.

Chang, J. T. & Hartigan, J. A. (1992), *Reconstruction of evolutionary trees from pairwise distributions on current species*, Citeseer.

Felsenstein, J. (1981), 'Evolutionary trees from DNA sequences: a maximum likelihood approach', *Journal of Molecular Evolution* **17**(6), 368–376.

Liò, P. & Goldman, N. (1998), 'Models of molecular evolution and phylogeny', *Genome Research* **8**(12), 1233–1244.

Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W. R., Hannon, G. J. & Smith, A. D. (2011), 'Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates', *Cell* **146**(6), 1029–1041.

Pupko, T., Pe, I., Shamir, R. & Graur, D. (2000), 'A fast algorithm for joint reconstruction of ancestral amino acid sequences', *Molecular Biology and Evolution* **17**(6), 890–896.

Steel, M., Hendy, M. D. & Penny, D. (1995), *Reconstructing evolutionary trees from nucleotide pattern probabilities*, Forschungsschwerpunkt Mathematisierung-Strukturbildungsprozesse, Univ.