Supplemental Text S1

1 The Evolutionary Model

Consider a set of τ *time periods* $\{t_j\}$ that are associated with tree edges. We assume perfect topological compatibility of all individual gene trees; henceforth we will identify these time periods with tree edges and completely ignore the topological information.

A gene g_i tends to evolve at an intrinsic rate r_i that is constant along time but deviates randomly along the time periods Let $r_{i,j}$ be the *actual* (or *observed*) rate of gene i at period j. Then $r_{i,j} = r_i e^{\alpha_{i,j}}$ where $0 < e^{\alpha_{i,j}}$ is a multiplicative error factor. The number of mutations in gene g_i along period j is hence $\ell_{i,j} = r_{i,j} t_j$, commonly denoted as the *branch length* of gene g_i at period j. Throughout the text, we will use i to identify genes g_i and j for time periods t_j.

We now detail on the input to the problem. We have m disjoint pairs of species, corresponding to m cherries in the tree. Henceforth, as these pairs form cherries in the tree and hence the paths between any two pairs do not intersect, we will consider each cherry as a single edge of length equal to the sum of the two cherry edges. Additionally, there are n genes. Hence, the input to the problem is a partial $n \times m$ matrix $A = [a_{i,j}]$ where $a_{i,j}$ (if exists) is the length of the edge between pair j according to gene i. Now, we can associate the length of the edge between pair j with a period of time, and denote this time as t_j. Recall that by our formulation $\ell_{i,j} = r_{i,j} t_j = r_i e^{\alpha_{i,j}} t_j$

Our final goal is to find, given the input matrix A, the maximum likelihood (ML) values for the variables r_i and t_j for $1 \leq i \leq n$ and $1 \leq j \leq m$. For this purpose, we assume a statistical model for r_i by assuming that $\alpha_{i,j} \sim N(0, \sigma^2)$ or equivalently, that $e^{\alpha_{i,j}}$ is log normally distributed, $e^{\alpha_{i,j}} \sim \ln N(0, \sigma^2)$. This allows us to formulate the likelihood function for a given gene and a species pair.

In order to calculate $Pr(\ell_{i,j})$ we take the log:

$$\log \ell_{i,j} = \log(t_j r_i e^{\alpha_{i,j}}) = \log t_j + \log r_i + \alpha_{i,j}, \tag{1}$$

and now compute $\mu_{i,j}$, the expected length of $\log \ell_{i,j}$:

$$\begin{aligned} \mu_{i,j} &= E[\log \ell_{i,j}] \\ &= E[\log t_j + \log r_i + \alpha_{i,j}] \\ &= E[\log t_j] + E[\log r_i] + E[\alpha_{i,j}] \\ &= \log t_j + \log r_i \qquad (2) \end{aligned}$$

where the second equation stems from Equation (1), the third from linearity of expectation, and the last equation follows since $\alpha_{i,j} \sim N(0, \sigma^2)$. The likelihood of an edge is the probability of seeing that length, $\ell_{i,j}$, given the model parameters. Specifically:

$$\begin{aligned} L(\ell_{i,j}) &= Pr(\ell_{i,j}|t_j, r_i) \qquad (3) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\mu_{i,j} - \log \ell_{i,j})^2}{2\sigma^2}\right) \qquad (4) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\alpha_{i,j}^2}{2\sigma^2}\right) \qquad (5) \\ & \qquad (6) \end{aligned}$$

where the last equation follows from (2).

As our input is the matrix A with data for n genes and m time periods, we have:

$$L(A) = \prod_{1\leq i\leq n} \prod_{1\leq j\leq m} Pr(\ell_{i,j}|r_i, t_j). \tag{7}$$

As the likelihood and the log likelihood obtain their minimum and maximum at the same parameter values, it is customary to search in the log space.

$$\begin{aligned}
\log L(A) &= \log\left(\prod_{1\le i\le N}\prod_{1\le j\le m} Pr(\ell_{i,j}|r_i,t_j)\right)\\
&= \sum_{1\le i\le n}\sum_{1\le j\le m}\log\left(\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\alpha_{i,j}^2}{2\sigma^2}\right)\right)\\
&= \sum_{1\le i\le n}\sum_{1\le j\le m}\left(-\frac{1}{2}\log(\sigma^2 2\pi)-\frac{\alpha_{i,j}^2}{2\sigma^2}\right)\\
&= -\frac{n_e}{2}\log\sigma^2-\frac{n_e}{2}\log 2\pi-\frac{RSS}{2\sigma^2} \qquad (8)
\end{aligned}$$

where n_e is the number of non empty entries in A and RSS is the residual sum of squares defined:

$$RSS=\sum_{1\le i\le n}\sum_{1\le j\le m}\alpha_{i,j}^2. \qquad (9)$$

As the first two terms in the right hand side of (8) are constant for a given input, it follows that maximizing the log likelihood is equivalent to minimizing RSS. Let $\widehat{RSS}$ be the optimal (ML) RSS obtained under the given parameters (we detail on this in the sequel). Under the ML formulation, we set $\hat{\sigma}^2=\widehat{RSS}/n_e$ where $\hat{\sigma}^2$ is the ML value for σ^2. Hence we get

$$\log\left(\hat{L}(GT_i)_{1\le i\le N}\right)=-\frac{n_e}{2}\log\left(\frac{\widehat{RSS}}{n_e}\right)-\frac{n_e}{2}\log 2\pi-\frac{n_e}{2} \qquad (10)$$

2 Minimizing RSS

The RSS is a polynomial over the variables r_i and t_j where every monomial is of the form:

$$\alpha_{i,j}^2=(\log\ell_{i,j}-\log t_j-\log r_i)^2. \qquad (11)$$

In our case we have $n=100$ rates corresponding to 100 genes and $m=57$ time periods corresponding to 57 pairs of species. In order to find the critical points of the RSS, we find the gradient of the RSS, that is the partial derivative of the RSS WRT every variable. The critical points are the points in the $m+n$ spaces where all these partial derivatives simultaneously vanish [2].

Finding these points can be done either using either some numerical method or any symbolic algebra package to find this solution. We used *Sage* [1] for this purpose.

Recall that the rates and time periods are mutually confound. and hence there are infinitely many solutions to the system. Hence we chose to assign arbitrarily the value 1 to r_0 which yields a unique solution. The Sage code and the solution for the RSS function can be found in he supplementary material.

References

[1] William Stein and David Joyner. Sage: system for algebra and geometry experimentation. *SIGSAM Bull.*, 39(2):61–64, June 2005.

[2] Gilbert Strang. *Introduction to Linear Algebra, Second Edition*. Wellesley-Cambridge Press, 1993.