SUPPLEMENTAL MATERIAL FOR HUNT AND CARRANO (2010)

PALEONTOLOGICAL SOCIETY SHORTCOURSE

Likelihood Inference for Evolutionary Models

Likelihood is defined as the probability of obtaining the observed data, given a specified

model. Accordingly, we need to know the probability distribution of evolutionary

outcomes. There are two different ways to construct likelihoods for evolutionary

observations: (1) take the original trait values (z) and construct a new set of observations

that are independent, (2) consider all observations jointly, accounting for their mutual

covariances.

(1) Likelihood of Evolutionary Differences

Applying strategy (1) to ancestor – descendant sequences, one can look at the

evolutionary increment in trait values (Δz) between successive samples in the time-series.

It can be shown that all the models considered in this chapter produce trait differences

that are normally distributed, with means and variances that depend on the model

parameters and the ages of the samples (Hunt 2006). The log-likelihood function for any

normal variable is the natural log of its probability density function:

$$\log L(\Delta z) = -\frac{1}{2}\log(2\pi V) - \frac{(\Delta z - M)^2}{2V},$$

where M is the mean and V the variance of the distribution of Δz. Directional evolution,

random walks, and stasis differ in their predicted means and variances for evolutionary

differences between ancestors and descendants. The logic for all the models is similar

and detailed derivations are provided in (Hunt 2006), and so here we will just sketch out

the findings for one model, the random walk. Each observed increment is the sum of all

evolutionary steps between ancestor and descendant. Each step has a mean of zero (an

assumption of the random walk), and a variance equal to the step variance. If t time

increments separate ancestor and descendant, the mean of this sum is zero and the

variance of this sum is t times the step variance. This follows because the mean and

variance of a sum of independent random variables is equal to the sum of their means and

variances, respectively. Moreover, by the Central Limit Theorem, this sum is

approximately normally distributed as long as t is not too small. Thus, for the random

walk, $M = 0$ and $V = t\,\sigma^2_{step}$. Substituting these terms in the equation above yields the

log-likelihood function for a single ancestor – descendant evolutionary increment.

Because increments are independent, the log-likelihood of a set of increments in a

sequence is the sum of all the separate log-likelihoods. The maximum-likelihood

estimate of σ^2_{step} is that value which yields the highest log-likelihood over all

evolutionary transitions.

The above assumes that trait means are known without error, which is of course false. As

a result, the expected variance of an evolutionary change from ancestor to descendant is

increased by an amount equal to the sampling variance on estimating the difference

between the ancestral and descendant means (Hunt 2006). The same basic logic holds for

calculating the log-likelihood for phylogenetic independent contrasts under Brownian

motion, except the sampling error term is somewhat more complex because it must

include estimation error for the calculated node values (see Felsenstein 1985, 2004).

(2) Joint Likelihood across all Populations

The second strategy for fitting models to trait data from ancestor–descendant or

phylogenetic relationships is to consider all observations jointly. Under all the models

considered here, the joint distribution of trait values is expected to be multivariate

normal, with a mean vector and covariance matrix determined by the model parameters

and the pattern of relatedness. In addition to the model parameters described in the text,

this approach requires an extra parameter, z_0 that represents the trait value at the root of

the tree or the start of the sequence. The details of this approach are described elsewhere

(Garland and Ives 2000, Blomberg et al. 2003, O'Meara et al. 2006, Hunt 2008), and so

we will describe only one model, directional evolution (BM with a trend), as an example.

The expected variance in evolutionary outcome in terminal taxon i is equal to the time

elapsed between the root of the tree and taxon multiplied by the step variance, $t_i\,\sigma^2_{step}$, the

same as noted above. The expected covariance among terminal taxa is equal to the step

variance, multiplied by the shared path length between those species, $t_{ij}\,\sigma^2_{step}$. This is

equivalent to the time elapsed between the root and the most recent common ancestor of

taxa i and j. Terminal taxa that are very recently diverged share most of their path from

the root, and therefore have high covariance. Sometimes all the time terms are collected

into what is called the phylogenetic variance-covariance matrix, $\mathbf{C}$, in which case the

covariance among terminal tips can be represented as $\sigma^2_{step}\mathbf{C}$ (see Fig. S1). Sampling

error adds to the diagonal of this matrix, increasing the expected variance of outcomes, but does not affect covariances.
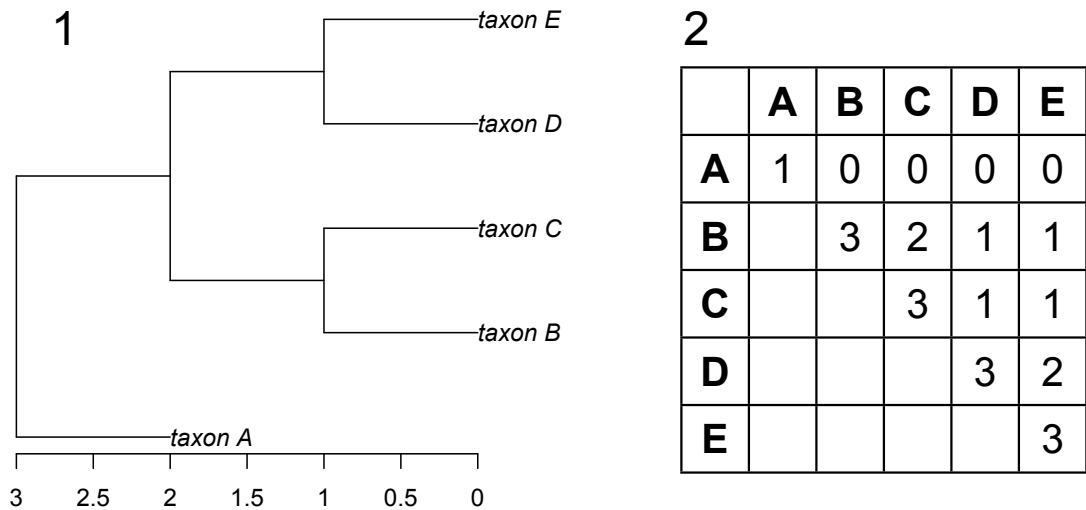
1

```
                      ─────────taxon E
         ┌───────────┤
         │           └─────────taxon D
  ┌──────┤
  │      │           ─────────taxon C
  │      └───────────┤
  │                  └─────────taxon B
──┤
  │
  └─────────────taxon A
```

| 3 | 2.5 | 2 | 1.5 | 1 | 0.5 | 0 |

2

	A	B	C	D	E
A	1	0	0	0	0
B		3	2	1	1
C			3	1	1
D				3	2
E					3

Figure S1.—*1*, example phylogeny of five taxa, with time axis on the bottom. *2*, the resulting phylogenetic variance-covariance matrix, **C**. Diagonal elements (variances) correspond to the time elapsed from the root to each terminal taxon. Off-diagonal elements (covariances) measure the amount of time between the root and the common ancestor of the two taxa in question.

For the trend model, the expected (mean) trait value increases linearly with the time, with a slope equal to the mean step: $m_i = z_0 + \mu_{step}\, t_i$. Calling this vector of means **m**, this means the probability of observing a vector of trait means across a sequence or tree is equal to the density function of the multivariate normal distribution with a mean **m** and

covariance matrix $\sigma^2_{step}\mathbf{C}$. Log-likelihoods are simply the natural logarithm of this density function.

The same logic applies to ancestor-descendant sequences, which can be thought of as completely pectinate trees with zero-length branches connecting each population to its ancestral node (Fig. S2). The models determine the nature of the multivariate mean vector and variance-covariance matrix, which are all that are required to compute the log-likelihood of a multivariate normal observation.
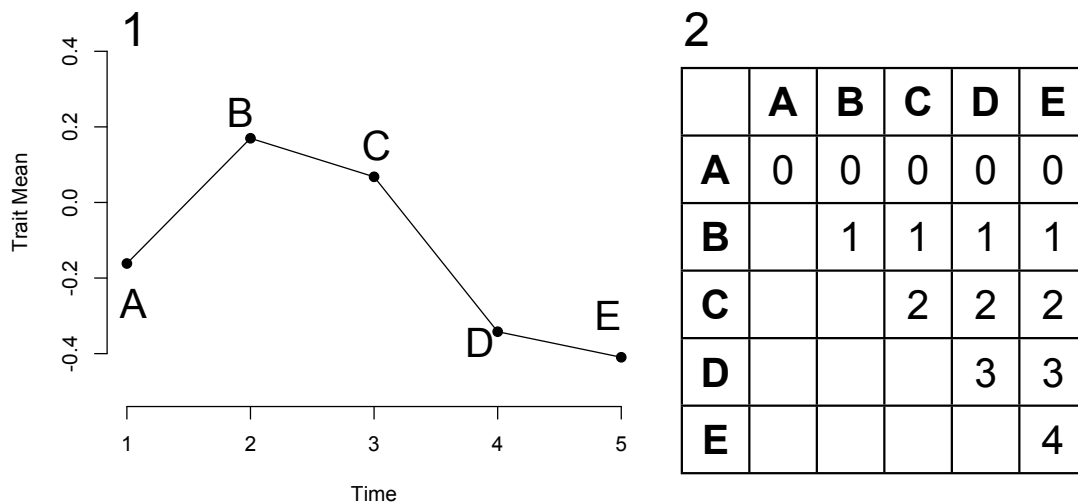

	A	B	C	D	E
A	0	0	0	0	0
B		1	1	1	1
C			2	2	2
D				3	3
E					4

Figure S2.—*1*, example time series with five populations (labeled A – E), with time axis on the bottom. *2*, the resulting variance-covariance matrix, **C**. Diagonal elements (variances) correspond to the time elapsed from the first population to each of the others. Off-diagonal elements (covariances) measure the amount of time between the first population (A) and the oldest of the two populations.

R objects: paleoTS and phylo classes

R allows creation of specialized data structures called classes. In this chapter, we have worked with two classes extensively: the `paleoTS` class for ancestor – descendant sequences of trait values, and the `phylo` class for representing phylogenies. The `paleoTS` class has four required elements. These are vectors of means (`mm`), variances (`vv`), sample sizes (`nn`) and ages (`tt`) of each sample in the sequence. The other elements of paleoTS objects are less important; see the help page for `as.paleoTS()` for details. Many of the functions in the *paleoTS* package were written to operate directly on `paleoTS` objects.

There are multiple ways of representing phylogenies as data objects. At present, the most common standard is the `phylo` object from the *ape* package. These objects are lists with several elements: an `edge` matrix that specifies the topology, a vector of `tip.label` for the names of terminal taxa, a vector of `edge.length` with branch lengths for each branch, and `Nnode`, which gives the number of internal nodes in the tree. The most important is the edge matrix that codes for the tree topology. This matrix has two columns and one row for each branch in the tree (internal and external). In this edge matrix, each of *N* terminal taxa is assigned a number from 1 to *N*, with internal nodes numbered consecutively thereafter, starting at the root. As an example, consider the tree in Figure S3.
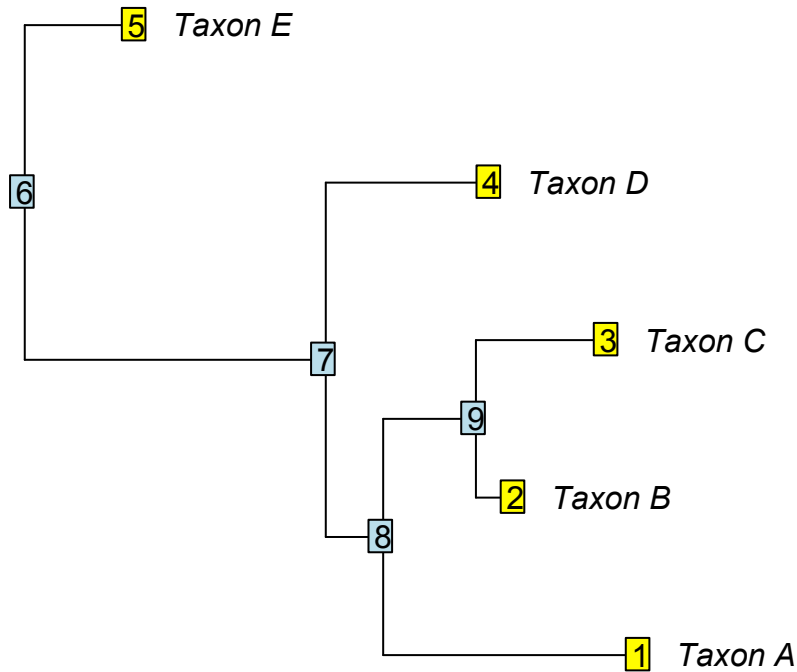
Figure S3.— Example tree used to demonstrate the `phylo` object class. The functions `tiplabels()` and `nodelabels()` from the *ape* package have been used to label the nodes and terminal taxa according to their numbers.

This tree has the following edge matrix:

```
      [,1]  [,2]
[1,]   6     7
[2,]   7     8
[3,]   8     1
[4,]   8     9
[5,]   9     2
[6,]   9     3
[7,]   7     4
[8,]   6     5
```

The first row represents the branch that connects node 6 (the root) to node 7; the last row connects node 6 to terminal taxon 5 (labeled Taxon E). For each branch, the left column is the ancestral taxon, and the right column is the descendant. It is informative to look at

the edge matrix to understand its structure, but users will almost never perform operations directly on it. Instead, there are a variety of helper functions that manipulate or use trees while hiding the underlying representation from the user (e.g., `drop.tip`, `extract.clade, plot, Ntip, Nnode, root, axisPhylo, tiplabels, nodelabels, multi2di, zoom`).

REFERENCES

Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioural traits are more labile. Evolution 57(4):717-745.

Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist 125(1):1-15.

Felsenstein, J. 2004. Inferring Phylogenies. Sinauer Associates, Inc., Sunderland, Massachusetts.

Garland, T., and A. R. Ives. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. American Naturalist 155(3):346-364.

Hunt, G. 2006. Fitting and comparing models of phyletic evolution: random walks and beyond. Paleobiology 32(4):578-601.

Hunt, G. 2008. Evolutionary patterns within fossil lineages: model-based assessment of modes, rates, punctuations and process. Pp. 117-131. *In* R. K. Bambach, and P. H. Kelley, eds. From Evolution to Geobiology: Research Questions Driving Paleontology at the Start of a New Century. The Paleontological Society.

O'Meara, B. C., C. c. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60(5):922-933.