

Modeling the Dynamics of Cultural Diversification

Cultural Phylogenies (Supplemental Tutorial)

Phylogenies have become an important metaphor and indispensable analytical tool for understanding the relatedness of organisms, but also languages, cultures, and artifacts. Broadly, a phylogenetic tree is a depiction of the inferred evolutionary relationships among a set of cultural lineages or taxa (i.e. species, genera, languages, cultural groups, stone tools, etc.). Here, we present a supplemental tutorial that explores the concepts and methods underlying phylogenetic inference and how they relate to the methods highlighted in the other tutorials.

In this tutorial, you will:

- Evaluate whether cultural phylogenies are possible
- Review how to read a phylogenetic tree
- Identify the necessary components for building a character-based phylogeny
- Explore the use of phylogenies to compare similarities between cultures by controlling for their shared history (i.e. Galton's Problem)
- Discuss potential alternatives to using phylogenetic approaches and when these alternatives should be considered

a. Are Cultural Phylogenies Possible?

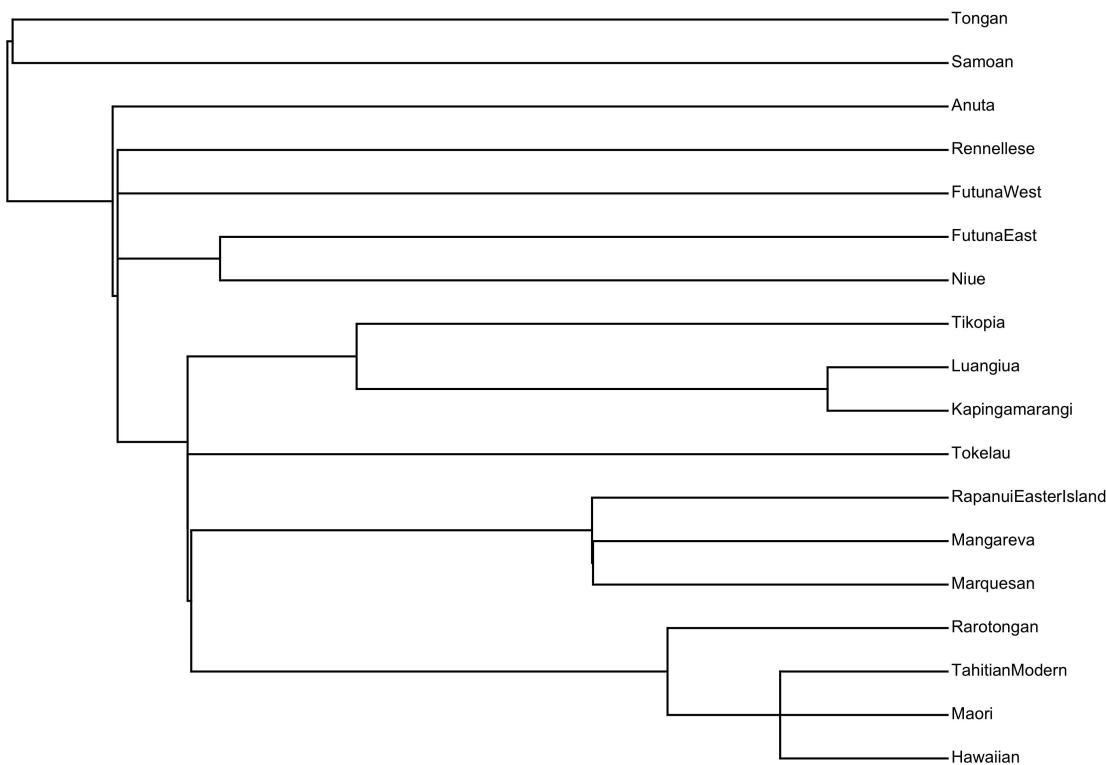
Models of cultural evolution posit that culture is information that is transmitted between individuals through various forms of imitation, teaching, and social learning (Boyd and Richerson 1985). This perspective views culture as having a second (or dual) system of inheritance that complements our genetic system of inheritance. This system of inheritance produces a hierarchy of cultural entities that are transmitted between individuals in the form of words, innovations, folk stories, religious beliefs, and social practices to name only a few ([Boyd et al. 1997](#)). Given this perspective on cultural inheritance, the first question to ask is whether the same assumptions and methods used to build phylogenies in biology can also be applied to human culture? In other words, are cultural phylogenies possible?

The applicability of cultural phylogenies is largely based on how one views the units of cultural change. If culture is passed from generation to generation (vertical transmission) and remains largely stable through time, cultural phylogenies are likely a valuable tool to examine historical relatedness. If culture is more likely to be passed from peer-to-peer (horizontal transmission) and is unstable through time, phylogenies may not be able to reliably infer relatedness in the past. To learn

more about how the units of cultural inheritance impact phylogenies, please watch the short video presented by Dr. Erik Gjesfjeld [here](#)

Check Your Understanding:

Regardless of the method used to construct a tree, how you read a tree remains the same. Below is a portion of the Austronesian language tree (discussed in detail below) with only languages that are associated with Remote Oceania / Polynesia. While many of you may be very familiar with reading trees, take a few moments to review your knowledge of trees by answering questions about the relationship between Polynesian ethno-linguistic groups.



1) The closest relative of Samoan is:

- a. FutunaWest
- b. Anuta
- c. Tongan
- d. Tikopia

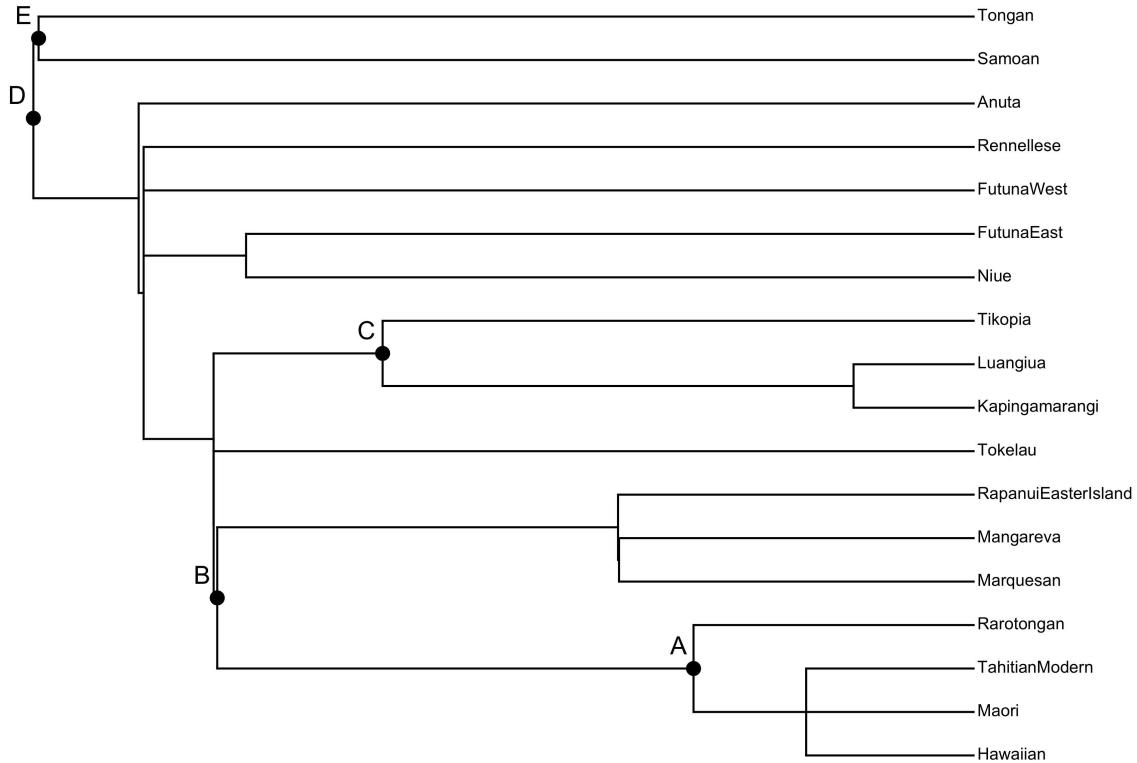
2) Which of the following is an accurate statement of relationships?

- a. Tikopia is more closely related to Niue than to Luangiuia
- b. Tikopia is more closely related to FutunaEast than to Kapingamarangi

- c. Tikopia is more closely related to Luangiuia than to Kapingamarangi
- d. Tikopia is equally related to Luangiuia and Kapingamarangi

3) Which of the following are most closely related to Anuta?

- a. Tongan, Samoan
- b. Tongan, Samoan, FutunaWest, Rennellese
- c. TahitianModern, Maori, Hawaiian
- d. It is equally related to all other languages except Tongan and Samoan



4) Using the tree with the nodes labeled, which of the five dots corresponds to the most recent common ancestor of a Marquesan and Maori?

- a. a
- b. b
- c. c
- d. d

5) Which of the five dots represents the most oldest divergence of languages?

- a. a
- b. b

c. c

d. d

► Answers

▼ b. Constructing Trees

Phylogenetic analysis can be broadly viewed as creating a hypothesis (or series of hypotheses) about historical relatedness. Most commonly, trees are created through the application of character-based methods. These methods aim to estimate the historical relationship between two taxa, such as cultural lineages, by inferring the pathways by which each evolved from a common ancestor. Using this approach, we require a few key ingredients in order to build a phylogenetic tree. These include:

1. Character data
2. Model of change
3. Optimality criterion

Character Data

Phylogenetic analysis can be performed with a wide variety of data including molecular sequence data, linguistic data, or technological data. However, all data used in a character-based phylogenetic analysis must have two components, the unit of analysis and character data about the unit of analysis. In cultural macroevolution, the unit of analysis is often a population of individuals or products that share common features. For example this could be group that shares a common language (English), a common ethnicity (Han Chinese), or even a shared technology (a particular house style). For each of these cultural units (or taxa), we also need to identify the variability within the character states. In some cases this may be as simple as a binary measurement of whether two groups share or do not share a similar word meaning or a cultural trait. In other cases, this might include multiple character states, such as series of projectile point measurements [O'Brien et al. 2002](#).

For example, in a recent chapter by Gjesfjeld and Jordan ([2019](#)), the authors use house styles from archaeological sites in the Great Plains of the United States as the cultural unit of analysis. At each archaeological site, at least one house style is characterized by a number of traits including the entrance direction, the number of hearths, or the pattern of wooden posts. In the table below, one house from each archaeological site is characterized by six characters and their multiple corresponding character state. For instance, the first character is the layout of interior posts (IntPost) with character states being 1 = Circular, 2 = Square / Rectangle, 3 = Linear, 4 = Double Linear.

Site (House)	IntPost	CentPost	ExtPost	NoExtPost	Entrance	Hearths	Size
Dodd (76)	3	1	3	8	4	2	4
Swanson (7)	4	3	3	6	2	3	1
Cannonball (4)	3	2	3	5	4	2	4
Shermer (2)	3	2	3	5	4	3	7
Molstad (4)	1	3	1	5	1	2	8

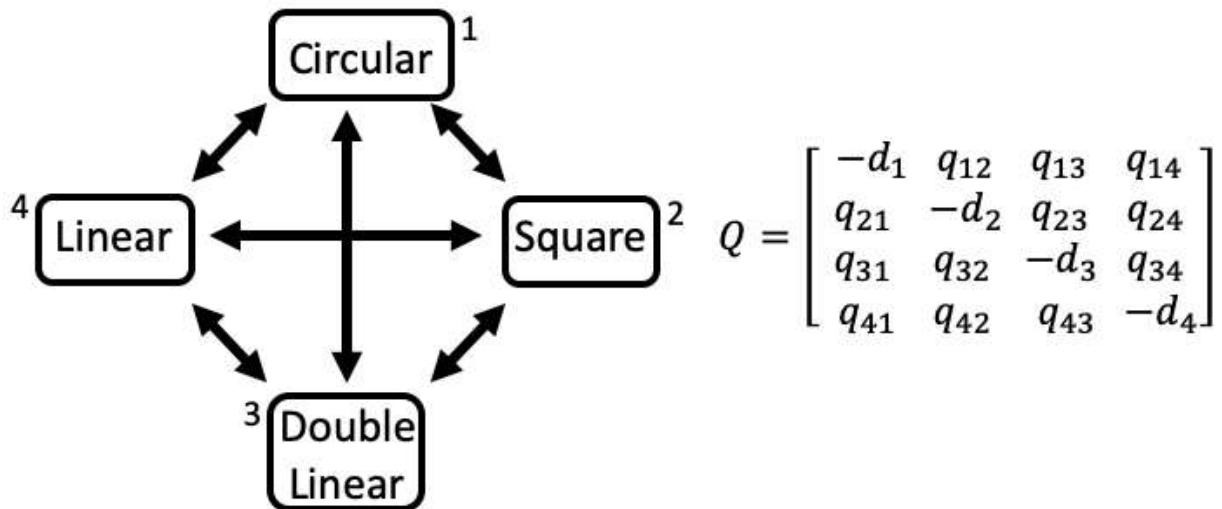
One of the central assumptions about data used for phylogenetic analysis is that it carries a phylogenetic signal. As discussed in the video lecture, this assumes that the cultural units used in the analysis are inherited through time (vertical transmission) and are relatively stable. This assumption may be more controversial for some sources of cultural data than others, as discussed below.

Model of Change

A majority of phylogenetic analyses currently rely on model-based methods to constructing trees. These methods use a mathematical description of evolutionary change within the characters, referred to as a "model". Broadly speaking, this model is the researcher's hypothesis about the mechanisms that influence evolutionary history.

In the house style data set presented above, the data can be considered as discrete, unordered data. As Harmon ([2018](#)) highlights, the most basic model for modeling changes in discrete characters is known as the Mk model. The model was originally developed by Pagel ([1994](#)), although the name Mk comes from Lewis ([2001](#)) and is a direct analogue of the Jukes-Cantor (JC) model used in molecular sequences ([Harmon 2018](#)).

One of the key assumptions of the Mk model is that changes from one character state to another only depends on the current state and not what came before ([Harmon 2018](#)). This is known as a Markov process and something that is discussed in the other tutorials of this module. The basic Mk Model also assumes that it is equally likely to change between any other states. If we use the house style data presented above, the transitions between character states for the first character would look something like this and can be summarized using a transition rate matrix:



It is important to note that the Mk model assumes that transitions among all possible character states can also occur at the same rate. However, this may not be a valid assumption for cultural or biological data. A range of other models do exist that fit more complex transition scenarios and will be discussed below.

Optimality Criterion

The earliest character based method for inferring phylogeny was the parsimony method ([Dunn 2015](#)). The parsimony assumption, often associated with a cladistic approach, can be viewed as the criterion for identifying which tree topology (e.g. the configuration and length of branches) is the most appropriate. The parsimony assumption aims to find the tree that minimizes the number of character changes required to produce the observed character states in the data. The use of a cladistic approach in cultural evolution has been particularly prevalent with archaeological data with extensive details about cladistics and archaeology found in O'Brien and Lyman (2003).

One potential issue in using parsimony methods is when there is potential for long branches (branches with lots of change). Often these branches might end up clustered together when they are actually quite distantly related ([Dunn 2015](#)). This is because the most parsimonious solution to the tree topology would be to place these branches closer to each other.

A more recent approach uses likelihood methods to estimate the tree topology. These methods aim to explain a set of data by quantifying how likely it was that the observed data was produced by a particular process ([Dunn 2015](#)). This process can be written as the likelihood (L) of seeing the data (D) under a particular hypothesis (H). The hypothesis (H) in a phylogenetic context is the chosen model of character change (e.g. the Mk model discussed above).

$$L = P(D|H)$$

Choosing the best tree when using likelihood methods proceeds by maximizing the likelihood of the parameters in the tree. These parameters are most likely to be branch lengths, transition values, etc. While calculating the likelihood is mathematically tractable for a tree with a small number of cultural units or taxa (languages, ethnic groups, technological products), it quickly become very difficult when trees get larger. For trees that have only 40 cultural units, there are more possible tree configurations ($1.0985 * 10^{57}$) than atoms in the entire world! ($1.33 * 10^{50}$)

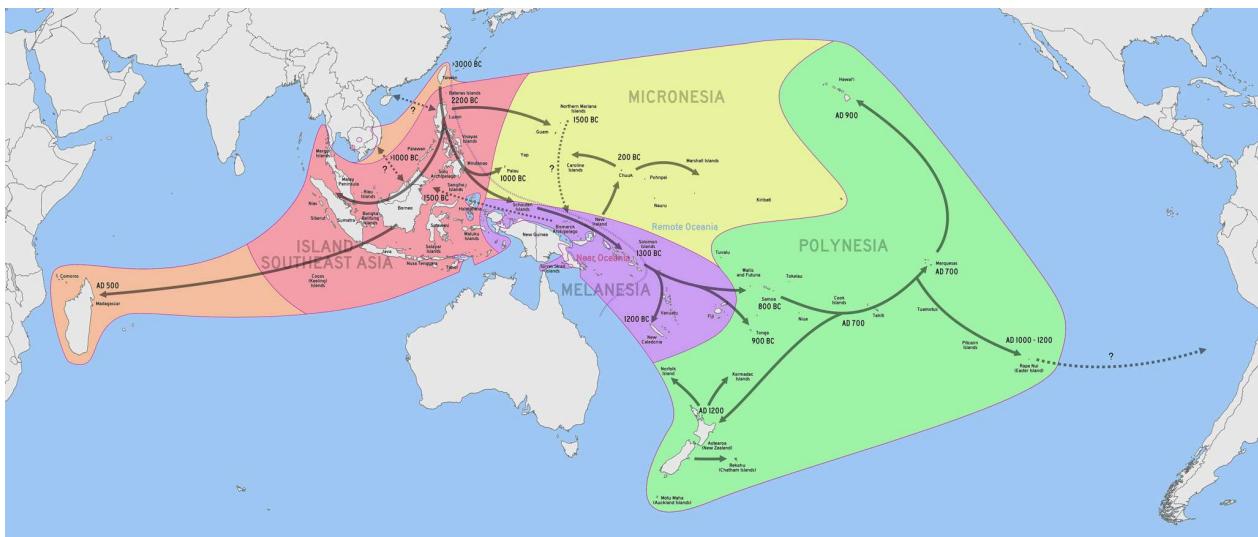
The most common solution to this exponential problem is to use Markov chain Monte Carlo (MCMC) to sample the possible trees that have the highest likelihoods. We talk about MCMC methods in other tutorials, but briefly just know that MCMC chains start at a random point in the space of all possible trees, moves through different tree configurations, and compares the likelihood of the current tree with the previous tree. If the new tree has a higher likelihood score, the MCMC chain moves towards it. This process continues millions of times as the MCMC chain slowly climbs towards trees with the highest likelihoods. To get a visual perspective on how MCMC methods search, visit the interactive gallery located [here](#).

Example: Austronesian Language Tree

One of the most well-known applications of a cultural phylogeny approach in Anthropology is associated with Austronesian-speaking peoples. If we use our list of key ingredients from above we can identify the following characteristics of this phylogeny.

1) **Cultural Data:** Languages are often considered to present a particularly suitable form of data for a phylogenetic approach as changes in language evolution often resemble the processes of biological evolution ([Gray and Jordan 2000](#)). In addition, languages are often closely linked to other cultural traits, contain large amounts of information, and evolve at rapid rates ([Gray, Drummond and Greenhill 2009](#)).

The Austronesian language family is one of the largest in the world and is spoken over vast geographical space. The dataset which underlies the Autronesian phylogeny is a large database of basic Austronesian vocabulary including words for animals, kinship terms, simple verbs, colors, and numbers ([Gray, Drummond and Greenhill 2009](#)). The words chosen to construct the phylogeny are those considered to be generally stable over time and less likely to be borrowed between languages (see video lecture above for greater details on core versus package traits). From this database, a set of homologous words, known as cognates, were identified for 400 well-known Austronesian languages ([Gray, Drummond and Greenhill 2009](#)). Broadly speaking, the unit of analysis for this tree is the ethno-linguistic group, meaning a population in which all individuals speak the same language and also have a shared history.



Distribution of Austronesian Language Speakers

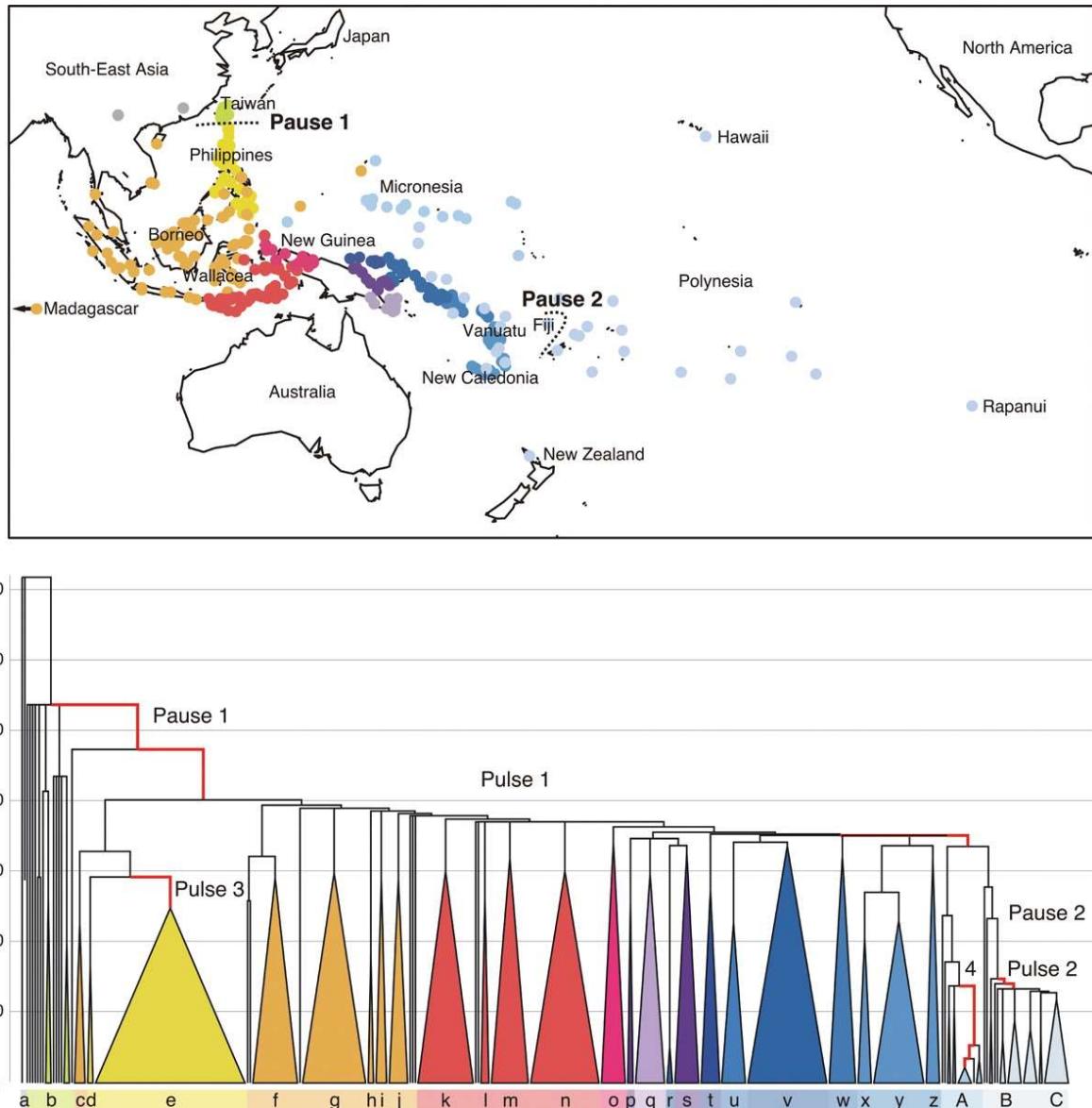
2) Model of Change As discussed above, most phylogenetic methods require a model of change that specifies how the changes between character states should be counted. In the modeling of language evolution, models of language change can range from simple to more complex depending on knowledge about the data. Here, a simple model would be a single parameter that represents the rate of change between the presence or absence of a cognate set (i.e. shared word meaning) (Greenhill and Gray 2009). We can also have a more complex and realistic model that would have a parameter for the ease of losing a cognate and a separate one for the ease of gaining a cognate. This might be useful as it is probably easier for a cognate set to be lost after it has arisen than for another ethno-linguistic group to independently derive the same cognate set. A model of language change that allows cognates to be lost many times, but gained only once is known as a stochastic Dollo model.

One of the advantages of model-based methods is that multiple models can be evaluated against each other to see which fits the data best. In the paper presented above by [Gray, Drummond and Greenhill \(2009\)](#), they use a two-state single-rate covarion model which can switch between fast and slow rates across the tree. More recent phylogenetic approaches can test between multiple models of character change and identify which model settings best fit the observed data (see Gjesfjeld and Jordan [2019](#)).

3) Optimality Criterion

Once an appropriate model of character change has been identified, it is necessary to identify the criterion with which to compare different tree topologies. In a Bayesian approach, we combine the likelihood highlighted above with our prior knowledge of trees to give us a posterior probability distribution of trees. Broadly speaking, the posterior distribution contains the trees that have the highest likelihoods and fit the data well given the data and priors (Greenhill and Gray 2009). In the creation of the Austronesian language phylogeny presented above, the MCMC was run for 100

million generations across six replicates with sampling of trees every 100,000 generations. The first 40 million generations (burn-in) ([Gray, Drummond and Greenhill 2009](#)) were discarded as the early trees sampled are strongly influenced by the starting values of the model. This left a total of 4,200 trees to use for further analysis. A summary of these 4,200 trees into a single tree is known as the maximum clade credibility tree and it looks like:



This figure can be found in Gray, Drummond and Greenhill ([2009](#)).

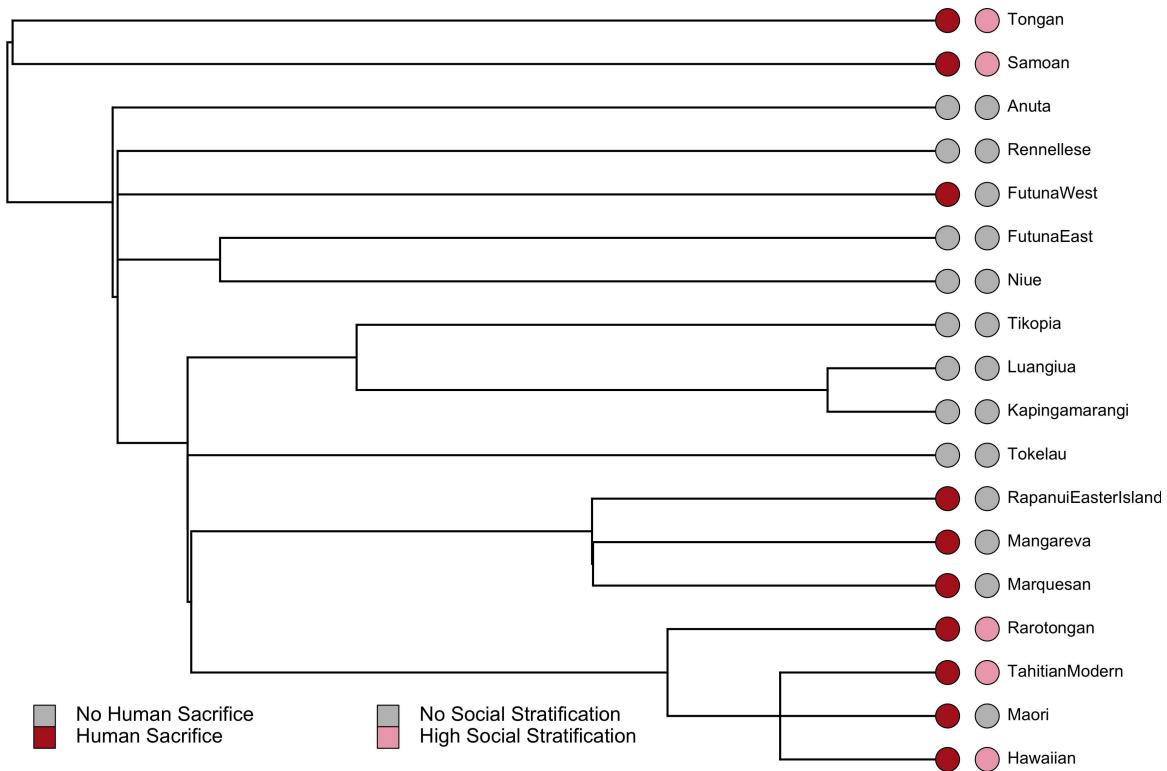
For those wanting further details about the programs, code, and other options for building language trees using Bayesian MCMC methods, we strongly recommend the excellent tutorial provided by Simon Greenhill and Konstantin Hoffman that can be found [here](#).

c. Phylogenetic Comparative Analysis

As highlighted above, phylogenies have a number of important roles in macroevolution. One of those is to estimate the relatedness between cultural units. However, knowing the historical relatedness between groups also helps us to solve another question, often known as **Galton's Problem**. The central issue of Galton's Problem is that cultural groups can not be treated as independent from each other when being comparing against each other (Mace and Pagel [1994](#)). This is because similar cultural groups might be similar because they have a shared ancestor and not because they independently developed the same traits. In general, phylogenies are advantageous to the comparison of cultures as they help control for *Galton's Problem* of auto-correlation. In other words, a cultural phylogeny helps to control for the influence of shared cultural history when comparing the similarities or differences between two or more cultural groups.

Watts et al. ([2016](#)) highlight the utility of a phylogenetic comparative approach by mapping the cultural traits of human sacrifice and social stratification onto a sample of the Austronesian Language tree. Broadly, the language tree in phylogenetic comparative methods is used as a proxy for the historical relatedness between different ethno-linguistic populations (i.e. Hawaiian, Tongan, Maori, etc.). The goal of these analyses is to view the presence or absence of specific cultural traits within the context of their historical relatedness. By doing this, we are able to evaluate whether two cultural groups are more likely to display similar traits because they share a more recent common ancestor.

Using a subest of only Remote Oceania ethno-linguistic groups, we can map on the presence or absence of specific traits, such as human sacrifice and high social stratification onto the phylogeny.



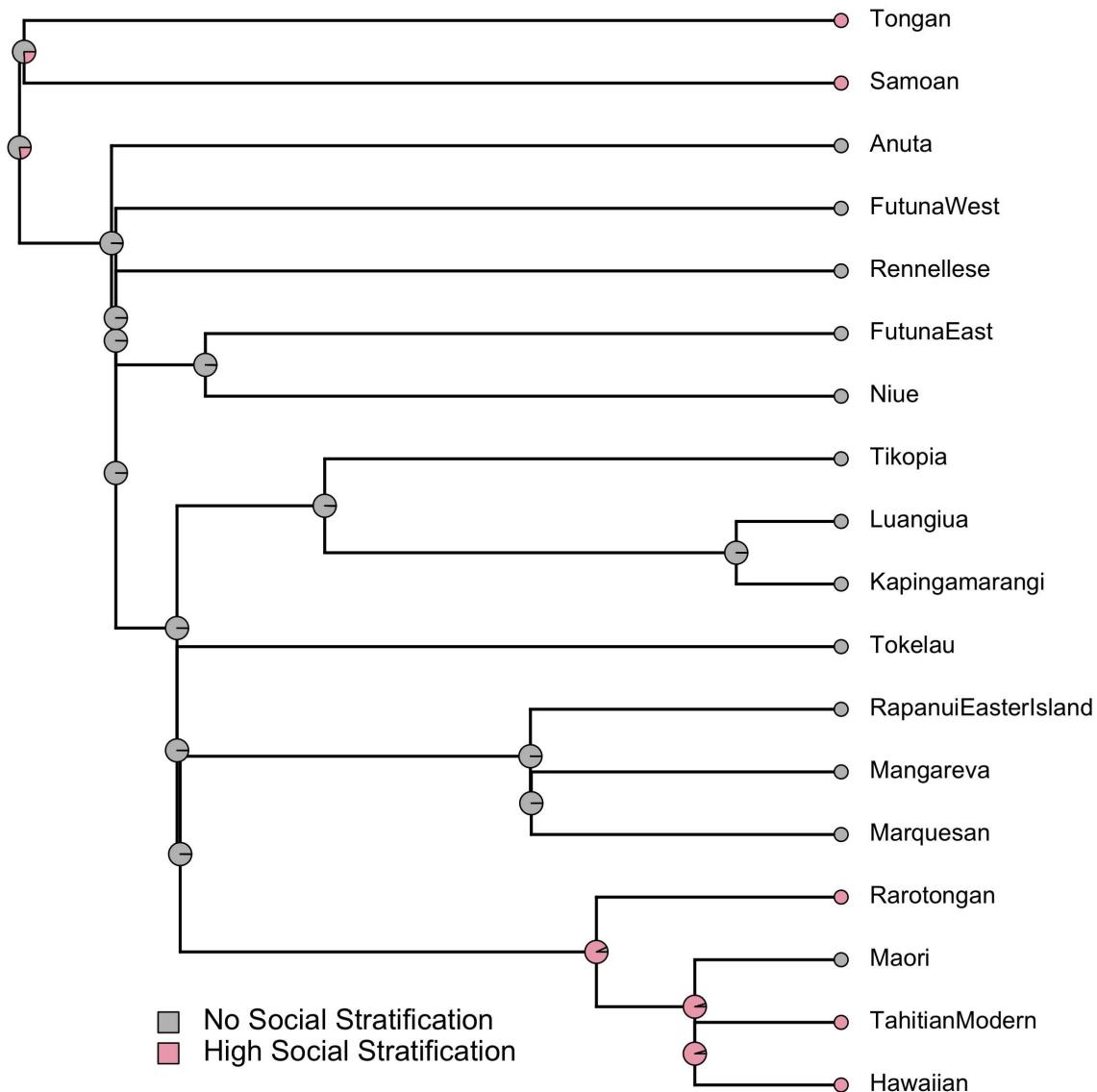
► R Code

One of the patterns that emerges is the presence of human sacrifice within the populations of Eastern Polynesian including Rapa Nui, Tahiti, New Zealand (Maori), and Hawaii. Most interesting is that numerous ethno-linguistic groups with social stratification also demonstrate human sacrifice. The correlation between the presence of these two traits leads Watts and his colleagues to suggest strong support for models in which human sacrifice stabilizes social stratification, once stratification has already arisen.

Ancestral State Reconstruction

Another contribution of phylogenetic comparative methods is their ability to estimate the traits of groups in the past, known as ancestral state reconstruction. Here this is done using a very simple method called 'ACE' from the `ape` package in R. This is provided here for demonstration purposes only and the reader is encouraged to explore more sophisticated and more appropriate methods for reconstructing ancestral states, such as those used by Watts et al. ([2016](#)).

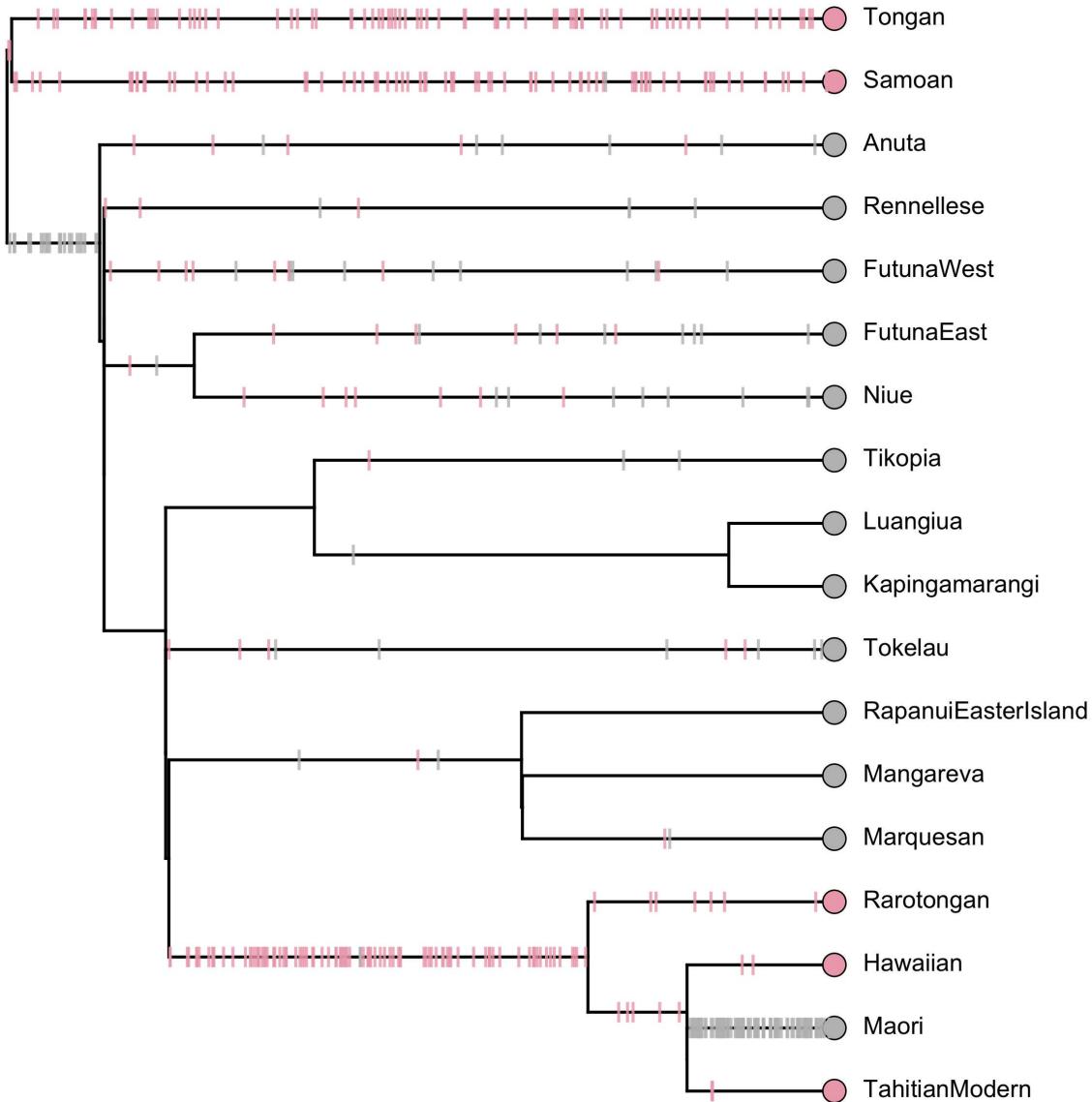
Broadly, we can see that most of the reconstructed ancestral states suggest little or no social stratification. The trait seems to appear only twice, in the ancestor of the Tongan and Samoan groups and in the ancestor of the Rarotongan, Hawaiian, and Tahitian groups. Maori represents an interesting example as it probably should demonstrate high social stratification given its ancestry, but it does not. This indicates that it likely lost this trait after the split from its related cultural groups.



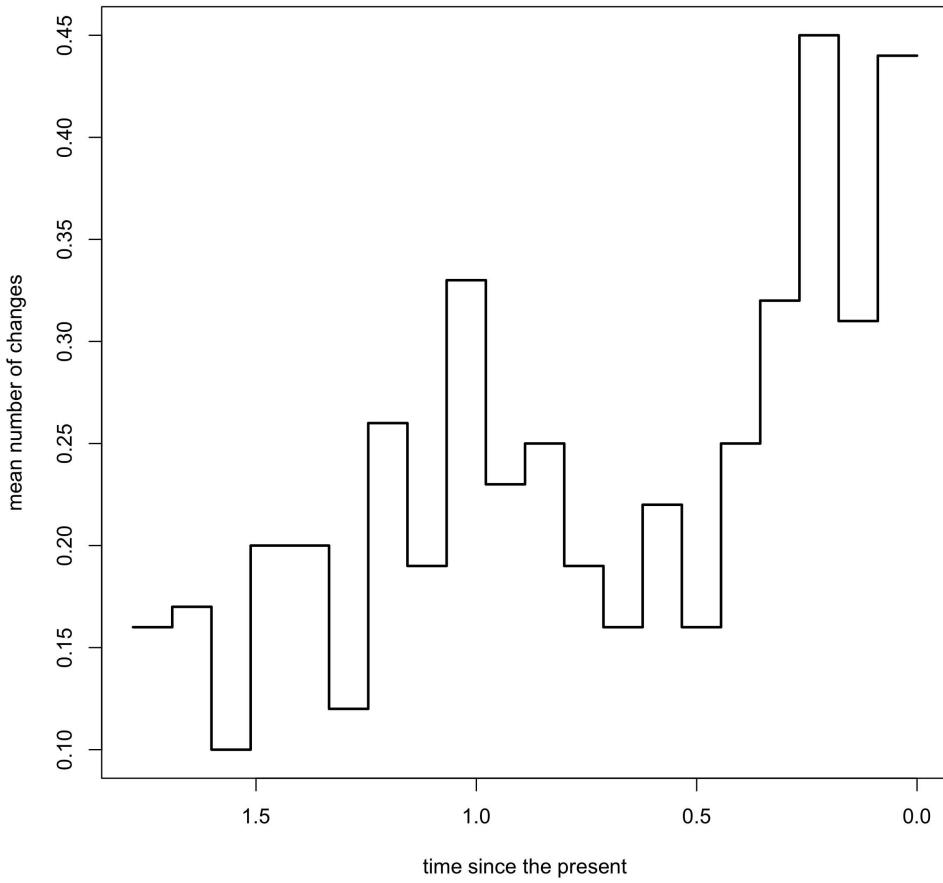
► R Code

Phylogenetic comparative methods provide a great way to solve Galton's Problem and compare cultures with each other. However, we may also be interested in other macroevolutionary questions such as the overall rate of character changes through time.

One relatively simple way to visualize change in a discrete character, such as social stratification, is highlighted by Liam Revell [here](#). This method uses a stochastic mapping procedure for reconstructing the history of a trait. Here, we use a simple equal rates (ER) model which assumes homogeneous rates of change across the tree. Broadly speaking, we are simulating 100 different character histories across the tree conditioned on transitions between characters in our data. We can then visualize a sample of character changes across 100 simulated histories. It is important to note that this procedure is not the same as fitting a model of heterogeneous rates through time, something that will we highlight in later tutorials.



If we continue to follow the phytools [tutorial](#) we can also compute the average number of changes through time. Here, we subset the tree into 20 different time segments and plot the average changes for each segment through time. The result generally shows an increase in the number of changes between 1,500 and 1,000 years ago and again from 500 years ago to the present. The sharp rise closer to the present is relatively unsurprising, as we would expect more changes to accumulate towards the present when we use an equal rates model. This is because the overall branch length becomes greater in those parts of the tree.



► [R Code](#)

d. Phylogeny Alternatives

Despite the ubiquity of phylogenetic approaches in studies of macroevolution, a tree-based approach may not always been the best option for certain types of data or macroevolutionary questions. For example, in their study of material culture artifacts, Témkin and Eldredge ([2007](#)) suggest that "Traditional phylogenetic analyses have rather limited application where traditional transmission is strong but intercultural exchange is weak". As indicated above, core vocabulary words are more likely to have strong vertical transmission and change less over time (i.e. cultural cores). However, traits associated with material culture or different technologies are potential candidates to change more quickly (i.e. cultural packages). Some studies have shown strong temporal continuity within material culture ([Tehrani and Collard 2002](#), [Lipo et al. 2006](#)), whereas other studies have also shown a tendency for strong horizontal transmission within material culture ([Témkin and Eldredge 2007](#), [Borgerhoff-Mulder et al. 2006](#)). Broadly speaking, the long-term and

vertical transmission of cultural traits is likely to differ significantly between cultural domains and should always be considered prior to the application of phylogenetic methods.

In addition to the concerns above, phylogenies can also be unreliable in circumstances when there is potential for a high degree of turnover and when a small number of cultural lineages persist to the current time period. These conditions could be met when examining technological or material culture, which can change quickly over time and as well as discarded. In these circumstances, archaeological or historical data on material culture can still be used to investigate changes in diversity through time, but a different conceptual and methodological approach needs to be taken.

In the remaining tutorials, we will introduce users to an alternative macroevolutionary framework that can examine the mode and tempo of cultural evolution using historical and material data. This approach complements phylogenetic methods, but does not require the same assumptions and therefore provides flexibility in the type of data that can be examined

Key Takeaways

- **Phylogenies are a valuable tool to inferring patterns of cultural macroevolution, especially when used with linguistic data.**
- **Phylogenies can be used in a comparative approach and help to disentangle whether two or more cultural groups share cultural traits, languages, or material culture due to their common ancestry.**
- **Phylogenies may be unreliable when applied to certain kinds of data, such as material culture. An alternative approach is to focus more directly on modeling changes in diversity through time using the temporal occurrences of cultural objects.**

References

Boyd, R., and P.J. Richerson. Culture and the Evolutionary Process. Chicago: University of Chicago Press, 1985.

Dunn, Michael. 'Language Phylogenies'. In The Routledge Handbook of Historical Linguistics, edited by Claire Bowern and Bethwyn Evans, 208–29. London: Routledge, 2015.

<https://doi.org/10.4324/9781315794013>.

Gjesfjeld, Erik, and Peter Jordan. 'Contributions of Bayesian Phylogenetics to Exploring Patterns of Macroevolution in Archaeological Data'. In *Handbook of Evolutionary Research in Archaeology*, edited by Anna Marie Prentiss, 161–82. Cham: Springer International Publishing, 2019.
https://doi.org/10.1007/978-3-030-11117-5_9.

Gray, Russell D., Alexei J. Drummond, and Simon J. Greenhill. 'Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement'. *Science* 323, no. 5913 (2009): 479–483.
<https://doi.org/10.1126/science.1166858>

Gray, Russell D., and Fiona M. Jordan. 'Language Trees Support the Express-Train Sequence of Austronesian Expansion'. *Nature* 405, no. 6790 (2000): 1052–1055.

<https://doi.org/10.1038/35016575>

Greenhill, Simon J., and Russell D. Gray. 'Austronesian Language Phylogenies: Myths and Misconceptions about Bayesian Computational Methods'. In *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*, edited by A. Adelaar and A. Pawley, 1–23. Pacific Linguistics, 2009.

Harmon, Luke. *Phylogenetic Comparative Methods · Learning from Trees*, 2018.
<https://lukejharmon.github.io/pcm/>

Lewis, Paul O. 'A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data'. *Systematic Biology* 50, no. 6 (2001): 913–925.
<https://doi.org/10.1080/106351501753462876>.

Lipo, C., M. Collard, M. O'Brien, and S. Shennan, eds. *Mapping Our Ancestors: Phylogenetic Approaches in Anthropology and Prehistory*. New Brunswick, NJ: Aldine Transactions, 2006.
<https://doi.org/10.4324/9780203786376>

Mace, Ruth, Mark Pagel, John R. Bowen, Keith F. Otterbein, Mark Ridley, Thomas Schweizer, and Eckart Voland. 'The Comparative Method in Anthropology [and Comments and Reply]'. *Current Anthropology* 35, no. 5 (1 December 1994): 549–64. <https://doi.org/10.1086/204317>.

Mulder, Monique Borgerhoff, Charles L. Nunn, and Mary C. Towner. 'Cultural Macroevolution and the Transmission of Traits'. *Evolutionary Anthropology: Issues, News, and Reviews* 15, no. 2 (2006): 52–64. <https://doi.org/10.1002/evan.20088>

O'Brien, M., and R.L. Lyman. *Cladistics and Archaeology*. Salt Lake City, UT: University of Utah Press, 2003.

O'Brien, Michael J., R. Lee Lyman, Youssef Saab, Elias Saab, John Darwent, and Daniel S. Glover. 'Two Issues in Archaeological Phylogenetics: Taxon Construction and Outgroup Selection'. *Journal of Theoretical Biology* 215, no. 2 (2002): 133–150. <https://doi.org/10.1006/jtbi.2002.2548>.

Pagel, Mark. 'Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters'. *Proc. R. Soc. Lond. B* 255 (1994): 37–45.

<https://doi.org/10.1098/rspb.1994.0006>.

Raup, David M., Stephen Jay Gould, Thomas J. M. Schopf, and Daniel S. Simberloff. 'Stochastic Models of Phylogeny and the Evolution of Diversity'. *The Journal of Geology* 81, no. 5 (1 September 1973): 525–42. <https://doi.org/10.1086/627905>.

Tehrani, J., and M. Collard. 'Investigating Cultural Evolution through Biological Phylogenetic Analyses of Turkmen Textiles'. *Journal of Anthropological Archaeology* 21, no. 4 (2000): 443–63. [https://doi.org/10.1016/S0278-4165\(02\)00002-8](https://doi.org/10.1016/S0278-4165(02)00002-8)

Tëmkin, Ilya, and Niles Eldredge. 'Phylogenetics and Material Cultural Evolution'. *Current Anthropology* 48, no. 1 (1 February 2007): 146–54. <https://doi.org/10.1086/510463>.

Watts, Joseph, Oliver Sheehan, Quentin D. Atkinson, Joseph Bulbulia, and Russell D. Gray. 'Ritual Human Sacrifice Promoted and Sustained the Evolution of Stratified Societies'. *Nature* 532, no. 7598 (April 2016): 228–31. <https://doi.org/10.1038/nature17159>.