

The Uses and Abuses of Tree Thinking in Cultural Evolution

Cara L. Evans^{1*}, Simon J. Greenhill^{1,2}, Joseph Watts^{3,4}, Johann-Mattis List¹, Carlos A. Botero⁵, Russell D. Gray^{1,6}, Kathryn R. Kirby^{1,7*}

¹Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Jena 07745, Germany; ²ARC Center of Excellence for the Dynamics of Language, ANU College of Asia and the Pacific, Australian National University, Canberra 2700, Australia; ³Religion Programme, University of Otago, Dunedin 9016, New Zealand; ⁴Centre for Research on Evolution, Belief and Behaviour, University of Otago, Dunedin 9016, New Zealand; ⁵Department of Biology, Washington University in St. Louis, St. Louis, MO 63130, USA; ⁶School of Psychology, University of Auckland, Auckland 1010, New Zealand; ⁷Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, M5S 3B2

*Authors for correspondence (caraeans1@gmail.com; kate.kirby@utoronto.ca).

Keywords: cultural evolution, phylogenetic comparative methods, cross-cultural research, cultural macro-evolution

Abstract

Modern phylogenetic methods are increasingly being used to address questions about macro-level patterns in cultural evolution. These methods can illuminate the unobservable histories of cultural traits and identify the evolutionary drivers of trait-change over time, but their application is not without pitfalls. Here we outline the current scope of research in cultural tree thinking, highlighting a toolkit of best practices to navigate and avoid the pitfalls and ‘abuses’ associated with their application. We emphasise two principles that support the appropriate application of phylogenetic methodologies in cross-cultural research: researchers should (1) draw on multiple lines of evidence when deciding if and which types of phylogenetic methods and models are suitable for their cross-cultural data, and (2) carefully consider how different cultural traits might have different evolutionary histories across space and time. When used appropriately phylogenetic methods can provide powerful insights into the processes of evolutionary change that have shaped the broad patterns of human history.

1. Introduction

Theories of cultural evolution are built on the observation that cultural features undergo innovation, modification and transmission. Over time, these processes have generated remarkable variation in human cultures. Humans speak around 7,000 distinct languages, affiliate with hundreds of religions, employ a range of kinship systems, engage in an array of subsistence practices, and adhere to a bewildering number of social conventions (Kirby et al. 2016). Phylogenetic methods provide a powerful approach to studying macro-evolutionary patterns of innovation, modification and transmission (Garamszegi, 2014; Mendoza Straffon, 2016; Gray & Watts, 2017). Their application to human culture has helped reinvigorate cross-cultural comparative research but has also been subject to criticism — both valid and misguided.

Phylogenies, also known as evolutionary trees, represent the common ancestry of populations and the splitting events that have occurred over the course of their history. Phylogenetic methods encompass a broad family of mathematical approaches that can be used to construct, analyse and incorporate phylogenies (Figure 1). Originally developed to study the evolution of biological organisms, these methods offer a general toolkit with the potential to provide answers to a range of cultural evolutionary questions.

An important distinction in cultural phylogenetics research is between methods of building trees (i.e., reconstructing the histories of cultural units based on assumptions of vertical transmission of cultural features (traits); Figure 1A), and methods that use previously-constructed trees in models that investigate the evolution and distribution of other cultural traits. A further important division in tree-thinking occurs between those methods and questions that simply detect and control for tree-like structure when examining variation in cross-cultural data (e.g., *What does the distribution of traits*

among societies tell us about the history of those societies and/or traits? Does horizontal or vertical transmission better explain the observed distribution of traits? Figure 1B), and those methods that require that the modelled data *are* tree-like (i.e. methods that ask: *What was the ancestral form of a cultural feature?* Figure 1C).

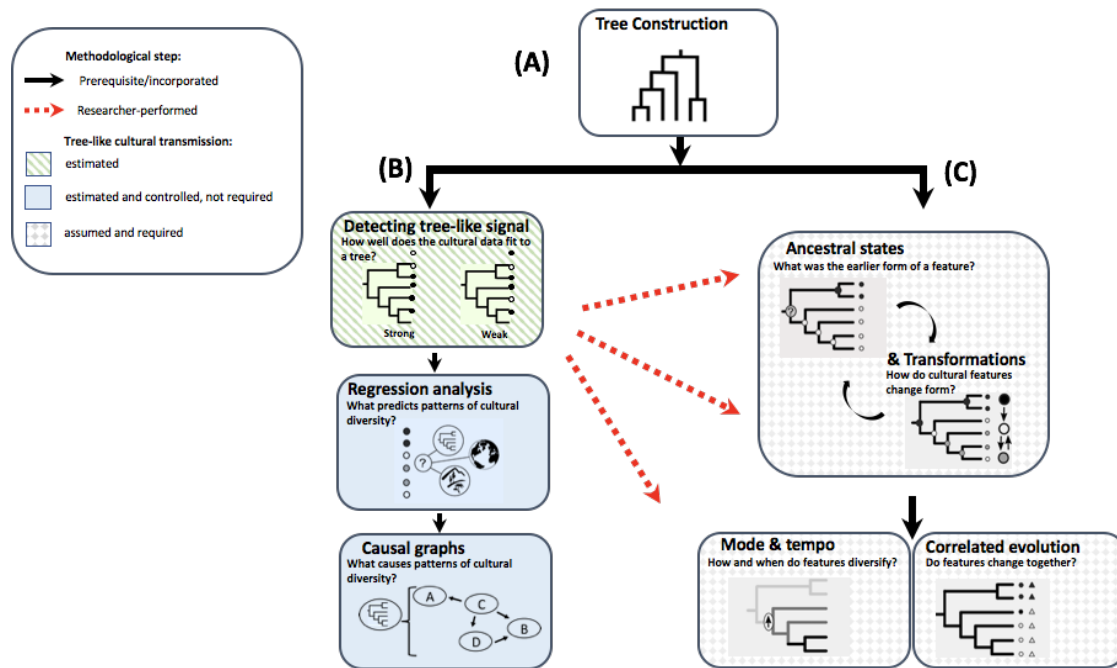


Figure 1. Phylogenetic methods that can be used to study cultural macroevolution. Black arrows indicate that the preceding methodological steps are directly incorporated in later methods: (A) Tree construction (Greenhill & Gray, 2009) is required for all subsequent steps; (B) testing for phylogenetic signal (e.g. Fritz & Purvis, 2010; Blomberg et al. 2003; Pagel, 1999) forms an integral part of phylogenetic regression (e.g. Atkinson et al. 2016; Minocher et al. 2019; Hadfield & Nakagawa, 2010), which in turn forms the basis of phylogenetic path analysis that can identify causal relationships; (C) Ancestral state reconstruction (e.g., Jordan et al, 2009), estimated in conjunction with rates of trait change and transformation (e.g., Currie & Mace, 2014; Currie et al. 2010), is required for models of trait correlation (Watts et al. 2016; Sheehan et al. 2018; Watts et al. 2015) and diversification (Gray et al. 2009; Cooney et al. 2017; but see Luca & Pennell, 2020). Red arrows indicate that suitable tests of phylogenetic signal (i.e., that the trait data fits sufficiently to the history inferred by the tree) should be conducted by the researcher before using methods detailed in (C; see also Section 2). Shading: grey shading indicates methods that both assume and require inferred historical relationships between the cultural units (tree taxa) to sufficiently reflect the history of the trait; green shading denotes methods that detect and quantify tree-like structure in cross-cultural data; blue shading denotes methods that detect and control for tree-like data structure among societies, but do not require it.

Phylogenetic methods offer exciting possibilities for a wide range of questions, only some of which explicitly require tree-like data. For data that *is* sufficiently tree-like, one of the strongest appeals of phylogenetic methods is that they offer the possibility to illuminate the unobservable past. Phylogenetic methods can reconstruct the ancestry of a vertically transmitted trait from the evolutionary signatures detected in its present-day distribution, even when archaeological records are entirely unavailable. However, despite this exciting potential, debate continues over how best to integrate cultural heterogeneity, disentangle the signatures of vertical transmission, horizontal diffusion and local socio-ecological drivers, and demonstrate that a cultural trait exhibits *enough* tree-like structure to justify using methods that reconstruct its evolutionary past (Fig. 1C).

Here, we review the application of phylogenetic methods in cross-cultural research. We focus specifically on the questions researchers should ask in order to avoid common methodological pitfalls when (1) deciding about the units of the underlying cultural data, (2) constructing trees, and (3) assuming tree-like transmission of other cultural features. Throughout, we outline a series of best practices and highlight emerging methods that promise to advance our understanding of macro-evolutionary patterns of mechanism and causation in culture.

2. Are the data appropriate for comparative phylogenetic analysis?

In the social sciences, phylogenetic methods have been used to build trees representing the evolution of a broad range of cultural units, including manuscripts (Howe et al. 2001), stone projectile points (O'Brien et al. 2001), textiles (Tehrani & Collard, 2002), languages (Grollemund, et al. 2015; Gray & Jordan, 2001; Gray et al. 2009; Sagart et al. 2019, Kolipakam et al. 2018), social systems (Matthews et al. 2013) and

nation states (Matthews et al. 2016). Cultural units – also referred to as taxa – are the entities represented at the tips of phylogenetic trees. A given cultural unit -- or taxon -- is typically compared to another based on its attributes ('features' or 'traits'). Below we outline three key considerations related to the structure, type and quality of the data used to delineate and describe cultural units and their traits. If ignored, the validity and reliability of inferences that can be drawn from the phylogenetic analyses described in Figure 1 – i.e., methods that construct trees and methods that use previously-constructed trees in models that investigate the evolution and distribution of other cultural features or traits – may be undermined by the data.

First, trait data should be structured around comparable units of analysis. This helps ensure variation in trait data reflects differences in the units' evolutionary histories, rather than differences in the scale of units being studied. For instance, the Database of Religious History contains descriptions of units varying in scale from single church communities, through religious families, to political empires (Slingerland & Sullivan, 2017). The evolutionary histories of the units, and the evolutionary processes acting on the units, are expected to vary across these different scales. Thus, researchers seeking to use phylogenetic methods to investigate these data should ensure their sample consists of religious units of a similar scale, and also that there is compatibility between the scale of the religious units and the units represented by the tree (see also Ember & Ember 2000; Slingerland et al. 2020).

Second, 'traits' must also represent comparable 'entities' across taxa (MacFarlane 2004). For example, cognate coding of word lists assumes concepts represented in the lists (e.g., bird, hand) were defined the same way across languages. Similarly, meaningful comparison of artifact morphology (e.g., projectile traits) requires that measurements across taxa were based on consistently identifiable measurement start- and end-points. In the case of cultural behaviours or practices, a trait (e.g., moralising high gods) must be defined in such a way as to be identifiable in very different contexts. Close attention to the

ways traits (concepts, morphological measurements, cultural practices) were defined for different taxa (languages, artifacts, societies) is critical to the quality of subsequent analyses.

These concerns are increasingly being addressed through the design of transparent and dynamic cross-linguistic and cross-cultural databases (Slingerland et al. 2020). This includes making available and linking detailed metadata to published datasets (e.g., detailed trait definitions, coders' notes on uncertainty, and links to primary sources), so that the definitions used when coding 'traits' are clear (List et al. 2019; Kirby et al. 2016). However, even complete with their metadata, coded datasets like Murdock's (1962-1971) *Ethnographic Atlas* are often limiting in that they delineate traits and alternative trait states (codes) based on the research interests and theories of the era in which they were built. In addition, they contain relatively little documentation on how coding decisions were reached, making it hard to evaluate the validity of the data (Purzicky & Watts, 2018). An alternative model is provided by the eHRAF Human Relation Area Files' *World Cultures* (Ember & Fischer, 2017), which provides users with finely indexed, searchable primary ethnographic materials.

Finally, variation - or conflict - of a given trait within a cultural unit, and variation in the focal dates of cross-cultural observations must be considered when selecting data for phylogenetic analysis (Atran, 2005). Variation in the expression of traits within a cultural unit is not commonly represented in cross-cultural datasets despite the potential to be widespread for many traits. For instance, societies can be assigned up to two different codes for a number of variables in the *Ethnographic Atlas* (Murdock 1962-1971), including 'exchanges at marriage,' 'post-marital residence location,' and 'house shape'.

Meanwhile, conflicting reports of the expression of a trait for a taxon are not uncommon and are often reported in coding notes (e.g., Murdock, 1962-1971), or may be represented in multiple, conflicting entries by different historical experts (Slingerland & Sullivan, 2017). Additionally, the difficulty of

obtaining synchronous data for multiple taxa can result in trait data for different taxa being based on observations collected over a span of several decades/centuries (Henderson & Whatley, 2014).

Such heterogeneities, contentions, and inconsistencies within the data sample can, of course, interfere with the accuracy of the inferences obtained from phylogenetic methods, which often assume single, unambiguous trait values at the tree tips for a given taxon, and sometimes require that trait measurements are synchronous in time (i.e., require, among other things, ultrametric trees). That said, an increasing number of comparative cultural studies are focusing on measures of variance (e.g., Borgerhoff Mulder et al. 2009) or measures of elasticities (or functions or associations; e.g., Ross et al. 2018). In phylogenetic analyses, an increasing number of solutions are being offered by Bayesian approaches that allow intra-taxon variation (Hadfield & Nakagawa, 2010) and ambiguity in the expression of traits (Meade & Pagel, 2019) to be incorporated in analyses (for more information on why Bayesian methods are the preferred method, and further information on choosing models and priors, see Greenhill et al. 2020; Maurits et al. 2020, Ritchie and Ho 2019). In addition to considering emerging methods for accounting for intra-taxon variation, we encourage researchers to consider the sensitivity of their inferences to both trait-measure inaccuracies and interference resulting from trait lability across varying sampling time-windows (see Watts et al. 2015 for an example). Together, these practices will improve the quality and reliability of results obtained from the application of phylogenetic methods to a given cultural dataset.

3. Tree construction: are phylogenetic trees accurate representations of cultural histories?

Phylogenetic methods have the potential to influence our understanding of the evolution of specific domains of culture (languages, artifacts) as well as of the histories of the populations (or other cultural

units) with which these cultural domains are associated. Phylogenetic methods can also provide a framework for cultural transmission against which alternative hypotheses about cultural change can be tested. Given this potential, how can we be sure that the trees resulting from phylogenetic analyses are accurate representations of cultural histories? In the sections below we outline a two-pronged approach to validate cultural phylogenies - combining simulation studies and careful benchmarking (e.g., Borgerhoff Mulder et al. 2006; Gray et al. 2007). We also emphasise the importance of using probability-based methods that estimate uncertainty in inferences about the tree topology.

Simulating trees: The first validation of tree construction methods is whether they can recover known trees after 'laboratory' manipulations have introduced 'noise' into the underlying data. A major concern with using trees as representations of history is that cultures can transmit information horizontally between groups through processes like cultural diffusion or linguistic borrowing (Moore 1994, Gray et al. 2010). One might expect that phylogenetic methods would break -- i.e., give the wrong result -- if there were traits in the underlying data that had been horizontally transmitted rather than inherited vertically from parent to daughter lineages. To test this, Greenhill et al. (2009) constructed a simulation study where they took two known phylogenies and used these trees to simulate datasets under varying levels of horizontal transmission, ranging from none up to a very high rate of 50% of all traits in the data being borrowed every 1000 years. Then for each of these simulated datasets they used Bayesian phylogenetic methods to reconstruct the phylogeny and compared the reconstructed phylogeny to the original known phylogenies. These results showed that Bayesian phylogenetic methods were, in fact, highly robust to borrowing -- able to correctly recover trees very similar to the original ones even under quite high levels of borrowing of ~15% every 1000 years.

However, this does not mean that borrowing is not a problem. Greenhill et al. (2009) also showed that the effect of borrowing is larger on unbalanced tree topologies with shorter branches, and when

borrowing is concentrated between a small set of branches. Further, studies investigating the effect of horizontal transmission on trait mapping have been less positive, with Nunn et al. (2010) arguing that their simulation results show that phylogenetic comparative methods should only be used when vertical transmission of traits was almost certain. However, a response by Currie et al. (2010), argues that the rates of horizontal trait transmission simulated by Nunn et al. were unrealistically high, and in fact the results from the phylogenetic comparative methods did not perform any worse than linear regression. Our recommendation is that caution is still needed when high rates of borrowing are to be expected, and that the effects of horizontal transmission are likely to interact with the tree shape and trait rates. Of course, with much higher levels of borrowing the initial pattern of vertical descent can become obscured. For example, in a phylogeny of football (Gray et al. 2007) geography trumped genealogy. Canadian football appeared closer to varieties of American football despite being historically derived from Rugby Union. The challenges of mechanistic (un)identifiability are discussed in more detail in section 4.

Benchmarking trees: Once constructed, trees should be benchmarked against other representations of the evolutionary history derived from alternate methods, data and approaches. Studies on a wide range of language families have generally found high concordance between the subgroupings identified by the traditional linguistic comparative method and Bayesian phylogenetic analysis of basic vocabulary: e.g. Sagart et al. 2019 (Sino-Tibetan), Kolipakam et al. 2018 (Dravidian), and Grollemund et al. 2015 (Bantu). For example, Gray et al. 2009 inferred the phylogeny of 400 Austronesian languages spoken in island South-East Asia and the Pacific. To validate their result they compared the subgroupings inferred on their trees to those proposed by traditional linguistic methods and found a very high concordance (e.g. Blust, 2009). Of the 400 languages, only 25 could be considered to be possibly misplaced and, critically, none of these misplacements disrupted the overall shape and pattern of the remaining languages (Greenhill et al. 2010, Greenhill & Gray, 2012; though see also Donohue et al. 2012). Furthermore, many

of the misplaced languages are involved in ongoing debates about their correct placement, indicating that the phylogeny is reflecting fundamental existing problems with placement of these languages (Greenhill et al. 2010). Most strikingly, the Austronesian basic vocabulary trees showed a close fit with both the expansion sequence and timing inferred from both archaeological and genetic data (Wilmshurst et al. 2011; Lipson et al. 2014). All three types of data support an initial origin in Taiwan approximately 5000 years ago, followed by a series expansion pulses and pauses across island South-East Asia and the Pacific. The basic vocabulary tree reflects this initial expansion across the Pacific even though there was subsequently a very substantial influx of people with “Papuan” genomes in regions such as Vanuatu (Posth et al 2018).

Quantifying uncertainty in the tree: Finally, the advent of Bayesian phylogenetic inference offers researchers the advantage of constructing a posterior distribution of possible phylogenies, whereby each unique tree topology and set of parameter settings is represented by its posterior probability. Bayesian inference sits in contrast to previous methods of phylogenetic reconstruction -- which instead produce point estimates of the single best tree -- allowing researchers to consider both the range of theoretically possible trees, along with the degree of ‘certainty’ with which the tree topology can be estimated (Cranston & Rannala, 2007, Rangel et al. 2015, Huelsenbeck et al. 2000). Moreover, many downstream methods that use the tree to investigate the evolution of other traits (Fig 1B-C), can use the posterior distribution of trees to integrate their inferences over uncertainty in the tree estimate.

4. Mapping other cultural features to lexical trees: divergent evolutionary histories, mechanistic (un)identifiability, and model shortcomings

We propose that the cultural trees that best match population history are generally those constructed from basic vocabulary. They typically exhibit relatively high levels of tree-like vertical transmission, robustness to realistic levels of borrowing, and match the archeological and historical records (see above). Once reliable linguistic trees are constructed (Figure 1A), they can be used to reconstruct the evolution of other cultural features (Figure 1C), to test and control for the phylogenetic non-independence of cultural units, and to examine alternative hypotheses about the forces underpinning phylogeographic distributions of particular cultural features (Figure 1B). However, in each of these cases there are potential pitfalls that should be avoided and which we explore in the following subsections.

4.1 Do different cultural traits have different evolutionary histories?

The extent to which basic vocabulary trees should be assumed to be good evolutionary models for other types of linguistic features (e.g., wider components of the lexicon or typological features), or for non-linguistic cultural traits (e.g. norms, rituals, subsistence and social structure) remains a topic of debate and a subject of empirical inquiry (Gray et al. 2010). The potential for trait-to-tree mismatch causes more serious methodological concerns for methods that require the cultural data to map to the same history as the tree (Figure 1C), than methods that estimate and control for tree-like non-independence between cultural units (Figure 1B).

Boyd et al. (1997) provide a useful framework for considering whether a given cultural trait, or set of traits, is likely to mirror the evolutionary history represented by basic vocabulary trees. At one extreme, Boyd et al. envision cultures as loose collections of ephemeral entities, dominated by horizontal transmission, and without integrated vertical transmission. At another, cultures are described as discrete entities that evolve as tightly integrated systems, as with vertebrate species (but not bacteria and viruses). A third possibility might be that cultures consist of core, vertically inherited traditions, but also contain some peripheral traits that are subject to much great horizontal transmission.

Finally, cultures could represent assemblages of coherent clusters, whereby phylogenetic methods might be appropriate on a cluster-by-cluster basis, but only if cluster boundaries and the relative importance of vertical vs. horizontal transmission within clusters, can be identified. Related to this scenario is the phenomenon of incomplete lineage sorting, well known in evolutionary biology as a source of historical incongruences. In biology, incomplete lineage assortment occurs when differences in genealogies are observed across some genetic loci and also with the overarching species tree. It results from incomplete sorting of an ancestral polymorphism during successive speciation events, leading to gene trees that coalesce at different points in time (Maddison & Knowles, 2006). Remarkably, for instance, a whole-genome analysis of the great apes has indicated that approximately 30% of our genome supports that humans split from chimpanzees earlier than they split from gorillas, or that these two species split from humans earlier than they split from each other (Hobolth et al. 2007). That gene trees are not always congruent with species trees highlights the inevitability that cultural trait phylogenies will not always display the same topology even when the underlying history is completely tree-like.

Incomplete lineage sorting might be quite common in language evolution, given the multiple sources of polymorphism that characterise language change (Jacques & List, 2019). There are two major sources for polymorphisms in language evolution -- sociolinguistic variation and linguistic variation. Neither of these two sources of variation can be easily filtered out during data preparation, which is why it is important to consider both when working with linguistic data. Sociolinguistic variation refers to the fact that a language is never used by a single person, but always by a larger group of people whose language may differ in various regards, even in their individual grammars (Dąbrowska 2020). Linguistic variation refers to variation in linguistic traits resulting in complex evolutionary scenarios which may be difficult to reconcile with a single tree. Since many linguistic traits represent a complex of smaller entities that tend

to evolve together, languages may even exhibit certain parallel changes long after their separation, a phenomenon commonly known as drift among linguists (Sapir 1921, List et al. 2016).

Figure 2 expands on an example by List (2016) informed by Kroonen (2013) to illustrate how the seemingly non-tree-like evolution of linguistic traits can be reconciled with an overall language phylogeny by invoking both sociolinguistic and linguistic variation. Here, linguistic variation is reflected in the fact that words in the Indo-European languages often have complex paradigms that were differently transmitted during the evolution of the Indo-European languages. As a result, a word like English “sun” reflects an ancient genitive (or more properly “oblique”) form, while a word like Spanish “sol” reflects an ancient nominative (“rectus”) form. Although both languages are related, they show different word forms for the concept “sun”, due to a specific form of linguistic variation that was inherited by many descendants of the Indo-European language. But linguistic variation is only one factor contributing to patterns that seem to contradict a given phylogeny. Another very common source of variation is sociolinguistic variation due to the fact that each language variety is spoken by many speakers at the same time who may well pronounce words differently or prefer certain words over other ones, in different contexts. According to Kroonen (2013) this must have been the case in Proto-Germanic, the ancestor language of Swedish and English, where two words for “sun” seem to have been in use at the same time, one which later became the ancestor of English *sun*, and one which later become the ancestor of Swedish *sol*. While the English form ultimately goes back to the old genitive and the Swedish form goes back to the old nominative paradigm form, the intermediate stage which can be reconstructed for Proto-Germanic was not a stage of one word showing two different case forms, but rather of two words of which we assume that they were used in different contexts or preferred by different speakers in the same language community.

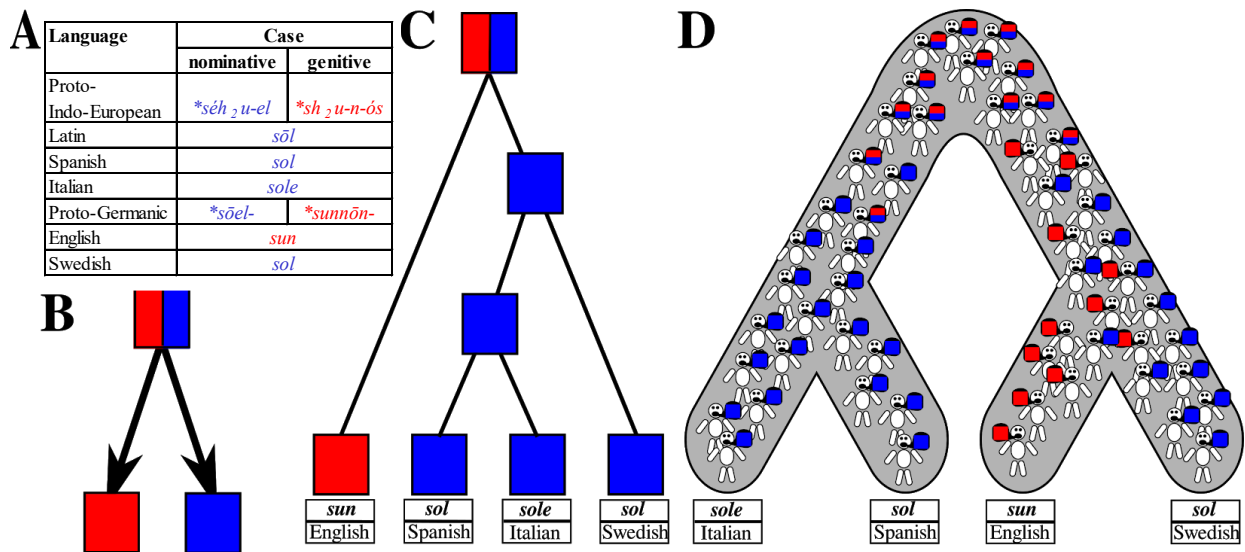


Figure 2. Incomplete lineage sorting in language evolution. The table in A shows words for sun in Indo-European languages along with intermediate stages, with two different forms for nominative and genitive case in Indo-European (similar to the irregular plural of mouse/mice in English), reflecting linguistic variation, and two independent word forms postulated for Proto-Germanic, reflecting sociolinguistic variation. B shows the basic linguistic change scenario by which either the nominative or the genitive form is preferred as a base form in complex paradigms. C shows a parsimonious but wrong language tree for extant languages, and D shows a reconciled scenario.

Importantly, using phylogenetic methods does not entail a commitment to the assumption that there is only one history. Indeed, phylogenetic methods can be used to investigate the extent to which different aspects of culture have congruent histories, and even to test alternative explanations for apparent incongruities. For example, Matthews et al. (2011) formally tested whether the motifs in pile-weave and non-pile-weave Iranian textiles shared the same history or not. They found that the two traditions were best explained by different phylogenies. Another study by Greenhill et al. (2017; See Figure 3) mapped lexical and grammatical data onto a phylogeny of Austronesian languages. So as to not bias the results in favor of the lexicon or grammar, the phylogeny was derived from independent historical linguistics research with subgroupings often primarily defined by phonological innovations, rather than grammatical or lexical features. They found stark differences in how these two subsystems of language

tracked the phylogeny: they were evolving at different rates (on average, the lexicon changed slower than the grammatical data), and the grammatical data showed a poorer fit to the phylogeny with more conflicting signal (homoplasy) probably caused by higher rates of horizontal diffusion. Greenhill et al. argue that these different subsystems of languages have differing dynamics that need to be carefully teased apart. Teasing these patterns apart is often only possible once you have a good estimate of the phylogeny. However, a recent study by Verkerk (2019) has applied a new 'multiple topologies' method (Pagel & Meade 2006) that simultaneously infers tree topology while assigning characters to these topologies. Methods like those in Matthews et al. (2011), Greenhill et al. (2017), and Verkerk (2019) provide a promising way forward for investigating the congruence of cultural histories without forcing all traits to share the same history.

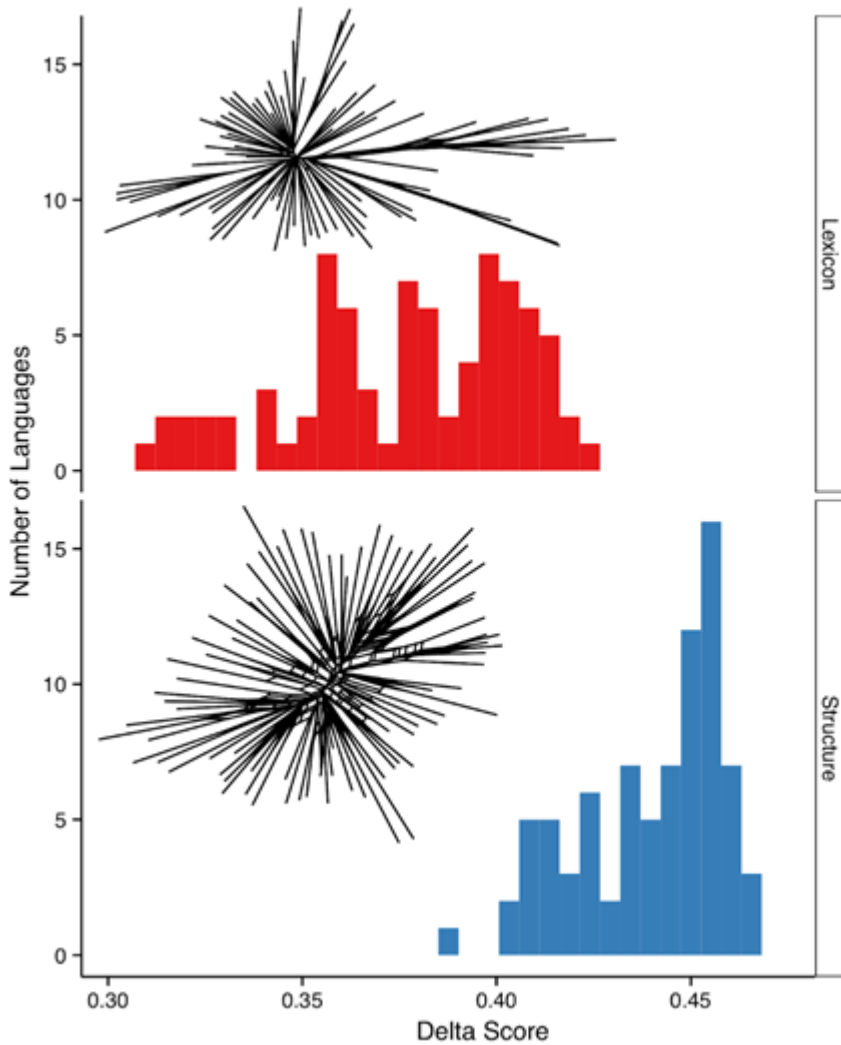


Figure 3: Histogram showing the phylogenetic fit of the lexical and grammatical/structural features of the languages from Greenhill et al. (2017) as measured by the Delta (δ) score (Holland et. al 2002, Gray et al. 2010). The grammatical/structural data show higher δ -scores, indicating a far worse fit to the overall phylogeny of Austronesian languages. The networks inset demonstrate the conflicting signal in these data visually, with the lexicon having a more treelike pattern with fewer conflicts. (Figure taken from Greenhill et al. 2017 with permission).

4.2 When is the use of methods that require historical coherence justified?

One subset of phylogenetic methods *does* require coherence and vertical transmission of cultural features (Figure 1C). Identifying *a priori* which features can be used in phylogenetic reconstructions is

often challenging, because of interference from processes such as horizontal transmission, incomplete lineage sorting, and independent invention (e.g., convergent evolution in ecologically similar environments). Debate over how best to demonstrate that a cultural trait exhibits *enough* tree-like structure to justify using methods that reconstruct its evolutionary past (Fig 1C) continues. We believe that the key to progressing against these difficulties is a three-step approach that considers (1) mode of transmission, (2) benchmarking practices against alternate lines of evidence, and (3) the continued development and careful utilisation of (i) methodological advances that disentangle pattern from process, and (ii) analytical approaches that can be applied when tree-like structure is missing or questionable in the data. Further discussion of the relationship between the histories of language phylogenies and cultural traits is provided in Slingerland et al (2020).

Modes of transmission and expression: First, it can be useful to consider what is already known about a trait's mode of transmission within and between groups, mode of expression (individual vs. group), and the extent to which a focal trait's expression and transmission can be disentangled from that of other traits (Borgerhoff Mulder et al. 2006; Towner et al. 2012). For instance, the present-day global distribution of a cuisine like pizza, which came into being in late 18th-century Naples, reveals much about the history of migration and economics, and relatively little about the cultural inheritance of food preferences. Similarly, the spread of major world religions might show relatively tree-like structure in their nested pattern of schisms, but the religion trees would reflect much more recent historical events than trees based on basic vocabulary. Attempting to reconstruct the history of Christianity on an Indo-European language tree would thus make no sense.

Benchmarking and validation: Currie et al. (2010) used phylogenetic comparative methods to infer the evolution of political complexity on the Austronesian tree and provided a good example of trait-history benchmarking and validation. Their results showed that, in the Pacific, political complexity generally

increased sequentially in small steps but that decreases in complexity could also happen through bigger drops (i.e. 'collapses'). Crucially, they were able to validate these inferences by comparing them to the historical, linguistic, and archaeological records. For example, the ancestral societal organisation of Proto-Oceanic was phylogenetically inferred to be limited to local communities, which is consistent with archaeological evidence showing only small-scale settlements (Kirch, 2000). Furthermore, archaeological and linguistic evidence also indicates that it is likely that this society had some form of social ranking (Hage 1999), which later became elaborated into simple chiefdoms such as those in the Trobriand islands and then further elaborated into complex chiefdoms in Polynesia like Tonga and Hawaii (Kirch & Green, 2000; Currie et al. 2010).

In another paper utilising the same language tree, Sheehan et al. (2018) investigate the coevolutionary relationship between intensive resource use and sociopolitical hierarchy, finding support for a reciprocal relationship between these two variables and highlighting the importance of both social and material factors as drivers of cultural complexity. Here the authors were able to validate their phylogenetic trait reconstructions with evidence from the archeological record: models that were constrained by the known history of intensive resource use were consistent with, and provided validation for, models that were given no constraints. In yet further analyses from the same region, Watts et al. (2015) highlighted that the relatively recent emergence of moralising high gods, as indicated by their ancestral state reconstructions, is consistent with early Muslim trade patterns in Southeast Asian cultures, and the concept of moralising high gods was 'borrowed in' to these societies during contact with traders. All the examples illustrate the fundamental role that historical benchmarking can play in validating cultural phylogenetic inferences, as well as the value of phylogenetic methods in contexts where horizontal transmission is important.

4.3 What about methods that detect yet do not require tree-like structure in the data?

Of course, many cultural traits do not leave any traces of their histories in the “fossil” record. Methods that detect and quantify the strength of tree-like structure in the data offer another line of validation, and do not assume tree-like evolution a priori. In fact, all traits being considered for phylogenetic reconstruction should first be formally examined for their potential tree-likeness (or phylogenetic signal) against the proposed tree model (see e.g., Watts et al. 2016). When a lack of tree-like signal is detected, and/or considerations of transmission mode or benchmarking imply that a trait is unsuitable for methods involving phylogenetic reconstruction (i.e., Figure 1C), phylogenetic methods that do not require tree-like data – but instead model the contemporaneous distribution of cross-cultural variation while controlling for detectable phylogenetic signal – can offer informative alternatives (i.e., Figure 1B). These latter methods might even be preferred, when a number of different traits and variables are being modelled, as phylogenetic reconstructions limit researchers to investigations of a maximum of two binary traits in coevolutionary models. Minocher et al. (2018), for example, using phylogenetic regression, found that marriage patterns in a globally-representative sample were best explained by pathogen-stress and male intra-sexual competition, after including eleven predictor variables, spanning a range of potential socio-ecological hypotheses, in their model. Exciting recent developments build further on this regression approach, allowing cross-cultural analyses to test hypotheses concerning directional causality in contemporaneous trait distributions - i.e., methods that detect shorter-term causal changes and do not require estimations of ancestral evolutionary states or processes- using phylogenetic path analysis (van der Bijl, 2018).

In the next two sections we will discuss how tests of phylogenetic signal can also be problematic as they sometimes falsely detect/generate the appearance of tree-likeness when alternate evolutionary

processes -- such as horizontal transmission, or independent invention in spatially or ecologically correlated environments -- are not appropriately modelled and controlled.

4.4 Are there correlations between the drivers of cross-cultural similarity that create a false impression of 'fit' to the language tree?

The extent to which phylogenetic methods can partition the different evolutionary drivers that have contributed to the present day phylogeographic distribution of a trait remains controversial. This involves teasing apart the relative importance of convergent evolution, horizontal transmission and vertical transmission. One of the greatest challenges for making inferences on the evolution of culture is that very different sources of cultural similarity can lead to virtually identical phylogenetic and geographic trait distributions in the present. For example, a trait may be spatially clustered because it is adaptive under a specific set of environmental conditions (habitat types are often spatially clustered themselves), because it was jointly inherited by a group of neighboring societies that descend from a common ancestor, or because it was more likely to diffuse horizontally among nearby groups that interact frequently with each other. Because neighboring groups are also typically close relatives and tend to inhabit similar habitats, distinguishing the effects of these correlated mechanisms can be extremely challenging. The continued development of methodological advances offers the most likely promise of resolution to this particular issue.

Methods to decouple ancestry, space, and ecology: Traditionally, cultural evolutionists have adjusted their sampling regimes to break down, at least in part, the correlation between diffusion, vertical transmission, and environmental selection in cross-cultural studies; hence removing the requirement to quantify or control for the effects of phylogeny. For example, some studies have explicitly sampled only geographically and phylogenetically distant societies (i.e. the Standard Cross-Cultural Sample (Murdock

& White, 1969)) to minimize the effects of horizontal and vertical transmission when testing hypotheses about the adaptive or social value of a trait (e.g., Ember et al. 2018). However, this stratified sampling is not completely effective -- the Standard Cross Cultural Sample does still have substantial and significant autocorrelation (Eff 2001), and it is better to tackle the problem head-on with phylogenetic methods rather than hoping that a stratified sample is sufficient (Bromham et al. 2018).

A number of recent approaches have instead actively sought to quantify the relative contributions of ancestry, diffusion and/or environmental selection on a variety of cross-cultural phenomena that are highly variable in their expression (e.g., Atkinson et al. 2016; Botero et al. 2014; Ember et al. 2020). These studies have relied on the analyses of large global datasets, using phylogenetic methods (i.e., phylogenetic regression; Figure 1B) that take advantage of the fact that neighboring societies are not always close relatives and/or do not always inhabit similar environments (i.e., hence allowing the methods to partial out the relative effects of each driver, e.g., Botero et al. 2014; Ember et al. 2020; Minocher et al. 2019).

In these studies, the potential for horizontal diffusion has often been assumed to be proportional to either the trait's representation within nearby cultural 'neighborhoods' (e.g., Botero et al. 2014) or the geographic proximity to neighboring groups (e.g., Atkinson et al. 2016). However, these proxies are likely to be problematic. Specifically, neighborhood proxies will often suggest a high potential for cultural diffusion if the trait of interest is spatially clustered, regardless of whether diffusion actually existed. Similarly, the use of centroid distances between cultural ranges, as typically defined in phylogenetic-spatial regressions, does not necessarily capture the actual amount of contact between neighbors. This problem arises because the size of cultural ranges tends to increase with latitude (Gavin & Stepp, 2014), meaning that centroid-to-centroid distances are likely to increase with latitude even if border-to-border distances do not. As a result, phylogenetic-spatial regressions based on centroid

distances could exhibit systematic underestimation of the potential for cultural diffusion as we move away from the tropics. A further dilemma is how to account for historical migrations/relocations of groups through (pre)history. In using spatial coordinates from one point in time to identify neighbouring cultures, current methods inevitably give precedence to neighbourhood effects from that time. Spatial neighbourhood approaches also do not account for long-distance borrowing among groups that may reflect contact through networks for resource extraction, trade, religion, and/or conquest.

Alternative approaches: Possible ways to more accurately disentangle the roles of space, environment, cultural diffusion and phylogeny without abandoning tree-based approaches might include approximating the potential for cultural diffusion by measuring the strength of connections in known contact networks, quantifying actual traveling times between known population centers using historically appropriate means of transportation, and estimating territorial overlap or shared boundaries rather than distances among centroids. However, the difficulties of modeling complex pathways of diffusion in cultural evolutionary studies has led some researchers to abandon altogether the idea that cultural relationships should be approximately tree-like and model them instead as reticulated networks (Bastide et al. 2018).

Phylogenetic network approaches were developed for biological entities where hybridization, horizontal gene transfer, and recombination are common. As such, they are also of great interest for the modelling of cultural and linguistic evolution. Unlike phylogenetic trees, phylogenetic networks allow for hybrid nodes (nodes with two parents). Phylogenetic networks can be unrooted, semi-rooted, or rooted. Rooted phylogenetic networks, like trees, can provide estimates of the timing of divergence, convergence (hybridization) and horizontal transfer of genetic (cultural) material (e.g., Solís-Lemus et al. 2017). Note that there is an important distinction between “implicit” or “data display” networks and “explicit” or “phylogenetic” networks (Morrison 2014, Nichols and Warnow 2008). Data display/implicit

networks are increasingly used to graphically represent conflict in phylogenetic trees (e.g., Fig. 3 in this paper). However, in these networks, internal nodes do not represent ancestors but noise or conflict in the tree signal. In contrast, internal nodes on explicit phylogenetic networks are meaningful such that hybrid nodes explicitly represent ancestral contact events.

In spite of their potential, explicit phylogenetic network approaches to inferring cultural phylogenies remain rare. In a recent review of their application to genomic data, Blair and Ané (2019) summarize three major obstacles to their broader application. First, network inference is computationally intensive. Inference based on ‘full network’ comparisons are currently only possible for small numbers of taxa (on the order of 5-10 taxa). Researchers working with larger samples of taxa could, in theory, run the network inference methods on multiple subsets of taxa. However, the researchers would then face the question of how to integrate topologies inferred for each subset, many of which may be unreliable simply for having been inferred from a very small sample (spurious findings of low to no phylogenetic signal are more common when examining trait distributions across a small number of taxa, and topologies inferred from small samples are more susceptible to distortion by ‘long-branch attraction’ (Heath et al. 2008; Roch et al. 2019; Blair and Ané 2019)). A second major obstacle is the paucity of methods for selecting the ‘best’ network from a series of inferred networks, particularly when some are more complex than others. One solution is to compare only a small number of plausible networks of similar complexity (e.g., Nakhleh et al. 2005). Indeed, methods for identifying cases where network approaches are appropriate, and for selecting a small set of ‘plausible’ networks for comparison, have been the focus of recent work in cultural evolution (e.g., Nakamura et al. 2020). A final obstacle to the adoption of network inference is its greater susceptibility, relative to tree inference, to violations of model assumptions (e.g., assumption of constant rates) which can lead to incorrectly inferred horizontal transmission events (Blair and Ané 2019).

In view of these obstacles, Blair and Ané (2019) argue it would be preemptive to abandon tree-based inference in favour of networks. Instead, they argue for the continued development of phylogenetic network approaches, and in particular of unbiased model selection methods that would allow researchers to weigh the suitability of alternative reticulated topologies. The authors also suggest fruitful paths for combining tree- and network-based inference methods that take advantage of the different strengths of the two approaches (e.g., Burbrink and Gehara 2018; Blair et al. 2019).

4.5 Are model shortcomings giving unwarranted precedence to tree-like inheritance patterns?

The correlated nature of the drivers of cross-cultural similarity creates yet another non-trivial issue for evolutionary analyses that attempt to disentangle them. Specifically, whenever a model “corrects” for one of these potential effects, it is implicitly assuming that such effect takes precedence over all others. For example, phylogenetic generalized least squares (PGLS) regression only tests the effects of potential ecological predictors after discounting cross-cultural similarities already expected from phylogenetic relatedness. In other words, these models implicitly assume that cultural similarity is first explained by descent, which as we know from some well-documented case studies may not always be the case. Take for example a trait that diffuses very quickly. Because neighboring cultures are often close relatives, a PGLS framework might incorrectly attribute observed spatial and phylogenetic clustering to vertical transmission and may lead interested researchers to miss the evidence for cultural diffusion altogether. A potential methodological solution to this problem is to assume that the baseline level of cross-cultural similarity is a joint function of phylogenetic and geographic distance. Recently implemented spatio-phylogenetic models (Dinnage et al. 2020; Freckleton & Jetz, 2009) use this approach and can estimate the most likely relative contribution of spatial and phylogenetic processes from the data themselves. Here too, though, it should be noted that the method is implicitly assuming a hierarchy of effects, in

which diffusion and vertical transmission take precedence over ecological selection (rendering the burden of proof for an effect of the latter much higher than for the former). Alternatively, when competing mechanisms are well-known and amenable to simulation, researchers may use generative inference (Kandler & Powell, 2018) to identify the most likely evolutionary mechanism. Generative inference involves comparing the patterns and or variable values observed in real data to the distribution of similar parameters obtained from simulations with alternative mechanistic scenarios.

5. Conclusion

One of the strongest appeals of cultural tree-thinking is that it offers a possible way to illuminate the unobservable past, and thus make causal inferences about the processes that have shaped human history. However, throughout this paper, we have cautioned that inferring processes from pattern requires careful consideration and validation. We would stress that cultural phylogenies should be treated as just one tool and one line of evidence. Other lines of evidence should also be explored, and we encourage researchers to consider the potential for multiple evolutionary processes (e.g., diffusion, ecological selection, cognitive constraints and descent), when comparing phylogenetic inferences with what is known from archaeology, anthropology and linguistics. As the field grows, new parallel data will make more robust benchmarking feasible. For this to happen datasets must be openly accessible and transparent regarding both their primary sources and coding definitions. The use and continued development of statistical methods that measure and disentangle phylogenetic signal from other drivers of cross-cultural diversity is also much needed. With careful use, phylogenetic methods can continue to play a crucial role uncovering the broad patterns of change in human cultural history and the processes that have shaped them.

Acknowledgments

CE, KK, SG, JML and RG were funded by the Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History. JW is supported by a grant from the Marsden Fund of New Zealand (19-VUW-188). JML was funded by the ERC Starting Grant 715618 “Computer-Assisted Language Comparison” (CALC, <https://digling.org/calc>).

References

- Atkinson, Q.D., Coomber, T., Passmore, S., Greenhill, S.J., Kushnick, G. 2016. Cultural and Environmental Predictors of Pre-European Deforestation on Pacific Islands. *PLoS One* 11, e0156340.
- Atran, S. 2005. Adaptationism for human cognition: Strong, spurious or weak? *Mind & Language*, 20(1): 39–67.
- Bastide, P., Solís-Lemus, C., Kriebel, R., William Sparks, K., Ané, C. 2018. Phylogenetic Comparative Methods on Phylogenetic Networks with Reticulations. *Systematic Biology*, 67, 800–820.
- Blair, C. & Ané, C. 2020. Phylogenetic Trees and Networks Can Serve as Powerful and Complementary Approaches for Analysis of Genomic Data. *Systematic Biology*, 69(3): 593–601, doi: [10.1093/sysbio/syz056](https://doi.org/10.1093/sysbio/syz056).
- Blair C., Bryson R.W., Linkem C.W., Lazcano D., Klicka J., McCormack J.E. 2019. Cryptic diversity in the Mexican highlands: thousands of UCE loci help illuminate phylogenetic relationships, species limits and divergence times of montane rattlesnakes (Viperidae: Crotalus). *Molecular Ecology. Resources*, 19:349–365. doi: [10.1111/1755-0998.12970](https://doi.org/10.1111/1755-0998.12970)
- Blomberg S.P., Garland T., Ives A.R. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile, *Evolution*, 57. 717-745.
- Blust R. 2009. *The Austronesian Languages*. Canberra: Pacific Linguistics
- Botero, C.A., Gardner, B., Kirby, K.R., Bulbulia, J., Gavin, M.C., Gray, R.D., 2014. The ecology of religious beliefs. *Proceedings of the National Academy of Sciences*, 111, 16784–16789.
- Borgerhoff Mulder, M., Nunn, C.L. and Towner, M.C. 2006. Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology*, 15(2), 52-64.
- Borgerhoff Mulder, M., Bowles, S., Hertz, T., Bell, A., Beise, J., Clark, G., Fazzio, I., Gurven, M., Hill, K., Hooper, P. L., Irons, W., Kaplan, H., Leonetti, D., Low, B., Marlowe, F., McElreath, R., Naidu, S., Nolin, D., Piraino, P., Quinlan, R., ... Wiessner, P. 2009. Intergenerational wealth transmission and the dynamics of

inequality in small-scale societies. *Science*, 326(5953), 682–688.
<https://doi.org/10.1126/science.1178336>

Bromham, L., Hua, X., Cardillo, M., Schneemann, H., & Greenhill, S. J. 2018. Parasites and politics: Why cross-cultural studies must control for relatedness, proximity and covariation. *Royal Society Open Science*, 5(8), 181100. <https://doi.org/10.1098/rsos.181100>

Burbrink F.T., & Gehara M. 2018. The biogeography of deep time phylogenetic reticulation. *Systematic Biology*, 67:743–755. <https://doi.org/10.1093/sysbio/syy019>

Cooney, C.R., Bright, J.A., Capp, E.J., Chira, A.M., Hughes, E.C., Moody, C.J., Nouri, L.O., Varley, Z.K. & Thomas, G.H. 2017. Mega-evolutionary dynamics of the adaptive radiation of birds. *Nature*, 542(7641), 344-347.

Cranston, K.A. & Rannala, B. 2007. Summarizing a posterior distribution of trees using agreement subtrees. *Systematic Biology*, 56(4), 578-590.

Currie, T.E., Greenhill, S.J., Gray, R.D., Hasegawa, T. & Mace, R. 2010. Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature*, 467(7317), 801-804.

Currie, T.E. & Mace, R. 2014. Evolution of cultural traits occurs at similar relative rates in different world regions. *Proceedings of the Royal Society B: Biological Sciences*, 281(1795), 20141622.

Dąbrowska, E. 2020. Language as a phenomenon of the third kind. *Cognitive Linguistics*, 31(2), 213-22.
DOI: <https://doi.org/10.1515/cog-2019-0029>

Dinnage, R., Skeels, A., & Cardillo, M. 2020. Spatiophylogenetic modelling of extinction risk reveals evolutionary distinctiveness and brief flowering period as threats in a hotspot plant genus. *Proceedings of the Royal Society B: Biological Sciences*, 287, 20192817.

Donohue, M., Denham, T. and Oppenheimer, S., 2012. New methodologies for historical linguistics?: Calibrating a lexicon-based methodology for diffusion vs. subgrouping. *Diachronica*, 29(4), pp.505-522.

Eff, E. A. 2001. Does Mr. Galton Still Have a Problem?: Autocorrelation in the Standard Cross-Cultural Sample. *World Cultures*, 15, 153–170.

Ember, M. and Ember, C. R. 2000. Testing theory and why the "units of analysis" problem is not a problem. *Ethnology*, 39(4):349-363.

Ember, C.R., Skoggard, I., Ringen, E.J., Farrer, M. 2018. Our better nature: Does resource stress predict beyond-household sharing? *Evolution and Human Behavior*. 39, 380–391.

Ember, C.R., Ringen, E.J., Dunnington, J. & Pitek, E. 2020. Resource stress and subsistence diversification across societies. *Nature Sustainability*, 1.1-9.

Ember, C.R., & Fischer, M.D. 2017. Using eHRAF World Cultures with other cross-cultural samples. In White, D. (Ed.) *Cross-cultural research companion*. NY: Wiley-Blackwell. 1-26.

- Freckleton, R.P., & Jetz, W. 2009. Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. *Proceedings of the Royal Society B: Biological Sciences*, 276, 21–30.
- Fritz, S.A. & Purvis, A., 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology*, 24(4), pp.1042-1051.
- Garamszegi, L.Z. ed., 2014. *Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice*. Springer-Verlag: Berlin.
- Giulano, P. & Nunn, N. 2018. Ancestral Characteristics of Modern Populations. *Economic History of Developing Regions*, 33:1, 1-17.
- Gavin, M.C. & Stepp, J.R., 2014. Rapoport's Rule Revisited: Geographical Distributions of Human Languages. *PLoS One*, 9(9), p.e107623.
- Gray, R.D., Bryant, D. & Greenhill, S.J., 2010. On the shape and fabric of human history. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1559), pp.3923-3933.
- Gray, R. D., Drummond, A. J., & Greenhill, S. J. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*, 323(5913), 479-483.
- Gray, R.D., Greenhill, S.J. & Ross, R.M. 2007. The pleasures and perils of Darwinizing culture (with phylogenies). *Biological Theory*, 2(4), 360-375.
- Gray, R. D., & Jordan, F. M. 2001. Language trees support the express-train sequence of Austronesian expansion. *Nature*, 409(6821), 743-743.
- Gray, R.D. and Watts, J., 2017. Cultural macroevolution matters. *Proceedings of the National Academy of Sciences*, 114(30), pp.7846-7852.
- Greenhill, S.J., Currie, T.E. & Gray, R.D. 2009. Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society B: Biological Sciences*, 276(1665), pp.2299-2306.
- Greenhill SJ, Drummond AJ & Gray RD. 2010. How accurate and robust are the phylogenetic estimates of Austronesian language relationships? *PLoS One*, 5(3): e9573.
- Greenhill, S.J. & Gray, R.D. 2009. Austronesian language phylogenies: myths and misconceptions about Bayesian computational methods. In Adelaar, A & Pawley, A. (Eds). *Austronesian historical linguistics and culture history: a festschrift for Robert Blust*. Pp. 375-397. Pacific Linguistics: Canberra.
- Greenhill, S.J. & Gray, R.D., 2012. Basic vocabulary and Bayesian phylolinguistics: Issues of understanding and representation. *Diachronica*, 29(4), pp.523-537.
- Greenhill, S.J., Wu, C.H., Hua, X., Dunn, M., Levinson, S.C. & Gray, R.D., 2017. Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences*, 114(42), E8822-E8829.
- Greenhill SJ, Heggarty P, & Gray RD. 2020 Bayesian Phylolinguistics. In Janda RD, Joseph BD, & Vance BS (Eds) *The Handbook of Historical Linguistics*, Volume II, pp. 226--253. Wiley-Blackwell: New Jersey

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C., & Pagel, M. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences*, 112(43), 13296-13301.

Hadfield, J.D., & Nakagawa, S. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies, and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, 23. 494-508.

Hage, P. 1999. Reconstructing Ancestral Oceanic Society. *Asian Perspectives*, 38, 200–228.

Heath T.A., Hedtke S.M., & Hillis D.M. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* 46:239–257. doi: 10.3724/SP.J.1002.2008.08016. Retrieved from: <https://www.jse.ac.cn/EN/10.3724/SP.J.1002.2008.08016>.

Henderson, M., Whatley, W. 2014. Pacification and Gender in Colonial Africa: Evidence from the Ethnographic Atlas. *Munich Personal RePEc Archive*. Retrieved from https://mpra.ub.uni-muenchen.de/61203/1/MPRA_paper_61203.pdf

Holden, C. J., & Mace, R. 2003. Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1532), 2425-2433.

Hopkins, M.J., Bapst, D.W., Simpson, C., & Warnock, R.C.M. 2018. The inseparability of sampling and time and its influence on attempts to unify the molecular and fossil records sampling and time in fossil and molecular records. *Paleobiology* 44, 561–574.

Howe, C. J., Barbrook, A. C., Spencer, M., Robinson, P., Bordalejo, B., & Mooney, L. R. 2001. Manuscript evolution. *Trends in Genetics*, 17(3), 147-152.

Huelsenbeck JP, Rannala B, & Masly JP. 2000. Accommodating phylogenetic uncertainty in evolutionary studies. *Science*, 288(5475):2349-50. doi: 10.1126/science.288.5475.2349.\

Jacques, G. & List, J.M. 2019. Save the trees. *Journal of Historical Linguistics*, 9(1), 128-166.

Jordan F.M., Gray R.D, Greenhill S.J, & Mace R. 2009. Matrilocal residence is ancestral in Austronesian societies. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 276: 1957–64

Kandler, A., & Powell, A., 2018. Generative inference for cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373, 20170056.

Kirby, K.R., Gray, R.D., Greenhill, S.J., Jordan, F.M., Gomes-Ng, S., Bibiko, H.J., Blasi, D.E., Botero, C.A., Bownern, C., Ember, C.R. & Leehr, D. 2016. D-PLACE: A global database of cultural, linguistic and environmental diversity. *PloS One*, 11(7), p.e0158391.

Kirch, P. V. 2000. *On the road of the winds. An Archeological history of the Pacific Islands before European contact*. Berkeley: University of California Press.

Kirch, P. V., & Green, R. C. 2001. *Hawaiki, Ancestral Polynesia: An essay in historical anthropology*. Cambridge University Press.

Kolipakam, V., Jordan, F.M., Dunn, M., Greenhill, S.J., Bouckaert, R., Gray, R.D. & Verkerk, A. 2018. A Bayesian phylogenetic study of the Dravidian language family. *Royal Society Open Science*, 5(3), p.171504

Kroonen, G. 2013. *Etymological dictionary of Proto-Germanic*. Brill: Leiden and Boston.

Lipson, M., Loh, P.R., Patterson, N., Moorjani, P., Ko, Y.C., Stoneking, M., Berger, B. & Reich, D., 2014. Reconstructing Austronesian population history in island Southeast Asia. *Nature Communications*, 5(1), 1-7.

List, J.M., Rzymiski, C., Greenhill, S., Schweikhard, N., Pianykh, K., & Forkel, R. (eds.) 2019. Concepticon 2.2. A Resource for the Linking of Concept Lists. Available online at <https://concepticon.clld.org>

List, J.-M., J. Pathmanathan, P. Lopez, and E. Baptiste (2016): Unity and disunity in evolutionary sciences: process-based analogies open common research avenues for biology and linguistics. *Biology Direct* 11.39. 1-17. DOI: <https://doi.org/10.1186/s13062-016-0145-2>

List, J.-M. (2016): Beyond cognacy: Historical relations between words and their implication for phylogenetic reconstruction. *Journal of Language Evolution*, 1.2. 119-136. DOI: <https://doi.org/10.1093/jole/lzw006>

Louca, S., & Pennell, M.W., 2020. Extant timetrees are consistent with a myriad of diversification histories. *Nature*, 580, 502–505.

Maddison, W.P. & Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, 55(1), pp.21-30.

Matthews, L.J., Edmonds, J., Wildman, W.J., & Nunn, C.L. 2013. Cultural inheritance or cultural diffusion of religious violence? A quantitative case study of the Radical Reformation. *Religion, Brain & Behavior*, 3(1), 3-15.

Matthews, L. J., Passmore, S., Richard, P. M., Gray, R. D., & Atkinson, Q. D. 2016. Shared cultural history as a predictor of political and economic changes among nation states. *PloS One*, 11(4).

Maurits, L., de Heer, M., Honkola, T., Dunn, M. and Vesakoski, O., 2020. Best practices in justifying calibrations for dating language families. *Journal of Language Evolution*, 5(1), pp.17-38.

MacFarlane, A. 2004. To contrast and compare. In Srivastava, V. K. (Ed.) *Methodology and Fieldwork*. Delhi: Oxford UP. 94-11.

Meade, A. & Pagel, M., 2019. BayesTraits, version 3.0. <http://www.evolution.rdg.ac.uk/BayesTraitsV3.0.2/BayesTraitsV3.0.2.html>

Mendoza Straffon, L (Ed). 2016. *Cultural Phylogenetics. Concepts and Applications in Archaeology*. Springer International Publishing: Berlin.

Meier-Brügger, M. 2002. *Indogermanische Sprachwissenschaft* [Indo-European Linguistics]. DeGruyter: Berlin.

- Minocher, R., Duda, P. and Jaeggi, A.V. 2019. Explaining marriage patterns in a globally representative sample through socio-ecology and population history: A Bayesian phylogenetic analysis using a new supertree. *Evolution and Human Behavior*, 40(2), 176-187.
- Moore, J. H. 1994. Putting Anthropology back together again: The ethnogenetic critique of cladistic theory. *American Anthropologist*, 96(4), 925–948.
- Morrison, D. A. 2014. Phylogenetic networks: A new form of multivariate data summary for data mining and exploratory data analysis. *WIREs Data Mining and Knowledge Discovery*, 4, 296–312. <https://doi.org/10.1002/widm.1130>
- Murdock, G.P., White, D.R. 1969. Standard Cross-Cultural Sample. *Ethnology*, 8, 329–369
- Murdock G.P. 1962-1971. Ethnographic Atlas, Installments I-XXVII. *Ethnology*, 1–10.
- Nakamura M, Wakano JY, Aoki K, Kobayashi Y. 2020. The popularity spectrum applied to a cross-cultural question. *Theoretical Population Biology* 133:104-116. <https://doi.org/10.1016/j.tpb.2019.10.003>
- Nakhleh, L. Ringe, D., & Warnow, T. 2005. Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages. *Language*, 81(2), 382-420.
- Nichols, J. & Warnow, T. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass*, 2(5): 760-820. Doi: 10.1111/j.1749-818X.2008.00082.x
- O'Brien, M.J., Buchanan, B., & Eren, M.I. 2016. Clovis colonization of Eastern North America: a phylogenetic approach. *STAR: Science & Technology of Archaeological Research*, 2(1), 67-89.
- O'Brien, M. J., Darwent, J., & Lyman, R. L. 2001. Cladistics is useful for reconstructing archaeological phylogenies: Palaeoindian points from the southeastern United States. *Journal of Archaeological Science*, 28(10), 1115-1136.
- Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature*, 401(6756). 877-884.
- Pagel, M & Meade, A. 2006a. Detecting Conflicting Phylogenetic Signal: A Mixture Model Approach with Multiple Tree Topologies. Paper presented at the New Zealand Phylogenetics Meeting, Kaikoura, New Zealand, February 12–17, 2006. Available at <https://zenodo.org/record/1117335#.WjeEzSOZPFR>
- Posth, C., Nägele, K., Colleran, H., Valentin, F., Bedford, S., Kami, K.W., Shing, R., Buckley, H., Kinaston, R., Walworth, M. and Clark, G.R. 2018. Language continuity despite population replacement in Remote Oceania. *Nature Ecology & Evolution*, 2(4), 731-740.
- Purzycki, B.G. & Watts, J. 2018. Reinvigorating the comparative, cooperative ethnographic sciences of religion. *Free Inquiry*, 38(3), pp.26-29.
- Rangel, T.F., Colwell, R.K., Graves, G.R., Fučíková, K., Rahbek, C. & Diniz-Filho, J.A.F. 2015. Phylogenetic uncertainty revisited: Implications for ecological analyses. *Evolution*, 69: 1301-1312. Doi: 10.1111/evo.12644

Ritchie, A. M. and Ho, S. Y. 2019. Influence of the tree prior and sampling scale on Bayesian phylogenetic estimates of the origin times of language families. *Journal of Language Evolution*, 4(2):108–123, <https://doi.org/10.1093/jole/lzz005>

Roch S., Nute M., & Warnow T. 2019. Long-branch attraction in species tree estimation: inconsistency of partitioned likelihood and topology-based summary methods. *Systematic Biology*, 68:281–297. Doi: 10.1093/sysbio/syy061

Ross, C. T., Borgerhoff Mulder, M., Oh S.-Y., Bowles, S., Beheim, B., Bunce, J., Caudell, M., Clark, G., Collieran, H., Cortez, C., Draper, P., Greaves, R. D., Gurven, M., Headland, T., Headland, J., Hill, K., Hewlett, B., Kaplan, H. S., Koster, J., Kramer, K., Marlowe, F., McElreath, R., Nolin, D., Quinlan, M., Quinlan, R., Revilla-Minaya, C., Scelza, B., Schacht, R., Shenk, M., Uehara, R., Voland, E., Willführ, K., Winterhalder, B. and Ziker, J. 2018. Greater wealth inequality, less polygyny: rethinking the polygyny threshold model. *Journal of The Royal Society Interface*, 1520180035. <https://doi.org/10.1098/rsif.2018.0035>.

Sagart, L., Jacques, G., Lai, Y., Ryder, R.J., Thouzeau, V., Greenhill, S.J. & List, J.M. 2019. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences*, 116(21), pp.10317-10322.

Sapir E. 1921. *Language. An Introduction to the Study of Speech*. New York:Harcourt, Brace.

Sheehan, O., Watts, J., Gray, R.D. and Atkinson, Q.D., 2018. Coevolution of landesque capital intensive agriculture and sociopolitical hierarchy. *Proceedings of the National Academy of Sciences*, 115(14), 3628-3633.

Slingerland, E., & Sullivan, B. 2017. Durkheim with data: the database of religious history. *Journal of the American Academy of Religion*, 85(2), 312-347.

Slingerland, E., Atkinson, Q. D., Ember, C. R., Sheehan, O., Muthukrishna, M., Bulbulia, J., & Gray, R. D. 2020. Coding culture: challenges and recommendations for comparative cultural databases. *Evolutionary Human Sciences*, 1-20.

Solís-Lemus C., Bastide P., & Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. *Molecular Biology & Evolution*, 34:3292–3298. Doi: 10.1093/molbev/msx235.

Tehrani, J. and Collard, M., 2002. Investigating cultural evolution through biological phylogenetic analyses of Turkmen textiles. *Journal of Anthropological Archaeology*, 21(4), 443-463.

Towner, M. C., Grote M. N., Venti J., Borgerhoff Mulder M. 2012. Cultural macroevolution on neighbor graphs: vertical and horizontal transmission among Western North American Indian societies. *Human Nature*. 23:283–305.

van der Bijl, W., 2018. phylopath: Easy phylogenetic path analysis in R. *PeerJ*, 6, e4718.

Verkerk, A., 2019. Detecting non-tree-like signal using multiple tree topologies. *Journal of Historical Linguistics*, 9(1), pp.9-69. Doi: [10.1075/jhl.17009.ver](https://doi.org/10.1075/jhl.17009.ver)

Watts, J., Greenhill, S.J., Atkinson, Q.D., Currie, T.E., Bulbulia, J. and Gray, R.D., 2015. Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proceedings of the Royal Society B: Biological Sciences*, 282(1804), p.20142556.

Watts, J., Sheehan, O., Greenhill, S.J., Gomes-Ng, S., Atkinson, Q.D., Bulbulia, J. and Gray, R.D. 2015. Pulu: Database of Austronesian supernatural beliefs and practices. *PLoS One*, 10(9), p.e0136783.

Watts, J., Sheehan, O., Atkinson, Q. D., Bulbulia, J., & Gray, R. D. 2016. Ritual human sacrifice promoted and sustained the evolution of stratified societies. *Nature*, 532(7598), 228-231.

Wilmshurst J.M., Hunt T.L., Lipo C.P., Anderson A.J. 2011. High-precision radiocarbon dating shows recent and rapid initial human colonization of East Polynesia. *Proceedings of the National Academy of Sciences*, 108: 1815–1820

Yang, Z. and Rannala, B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Molecular Biology & Evolution*, 14(7). 717-724.