



Meaning and Purpose: Using Phylogenies to Investigate Human History and Cultural Evolution

Lindell Bromham¹

Received: 19 February 2021 / Accepted: 1 March 2022 / Published online: 29 April 2022
© The Author(s) 2022

Abstract

Phylogenies are increasingly being used to investigate human history, diversification and cultural evolution. While using phylogenies in this way is not new, new modes of analysis are being applied to inferring history, reconstructing past states, and examining processes of change. Phylogenies have the advantage of providing a way of creating a continuous history of all current populations, and they make a large number of analyses and hypothesis tests possible even when other forms of historical information are patchy or nonexistent. In common with approaches taken in other historical sciences, phylogenetics is a way of reconstructing past and processes using the traces left in the present day. Trees, based on DNA, language, cultural traits, or other evidence, are now sprouting all over the academic landscape. The increasing use of phylogenetic analysis to understand human cultural evolution has been embraced by some, and scorned by others. The purpose of this article is not to review methods and applications of phylogenetic analyses, nor to consider the growing field of cultural phylogenetics, but, more broadly, to explore how we interpret phylogenies as narratives about human diversification. The first half of the article deals with meaning: phylogenies are often interpreted as histories, but a bifurcating tree is at best an abstract representation of history, and its connections to past events and processes is dependent on the data used, the assumptions made in the analysis, and the degree to which nodes in the tree (where one lineage splits into two) can be connected to change and movement in real populations. The second half of the article explores the purpose of phylogenies: a tree does not have to be a literal history of human lineages in order to be useful for investigating processes of human diversification. Phylogenies should not be read as accurate records of history, but as a way of exploring plausible explanations for current patterns of diversity. Phylogenies provide important information that can be used to test ideas about human diversity, and can help to guard against errors of inference arising from statistical artifacts.

Keywords Cultural evolution · Galton's problem · Language evolution · Phylogenetics · Phylogenetic nonindependence

Meaning: How Should We Interpret Phylogenies?

Phylogenies are increasingly being applied in human cultural evolution, to understand how human capacity for culture and language evolved (e.g., Lind et al. 2013), to reconstruct mechanisms and processes of cultural evolution (Mace and Holden 2005; Gray et al. 2007; Currie 2013), and to reconstruct the trajectories of change for particular

aspects of culture (e.g., Watts et al. 2015). We can distinguish two kinds of phylogenies that can be used to explore the processes of human cultural evolution, history, and diversification. Firstly, cultural phylogenies can be based on the products of cultural evolution itself, whether material objects generated by processes of cultural or technological evolution such as musical instruments (Temkin and Eldredge 2007), stone tools (O'Brien et al. 2014), woven textiles (Tehrani and Collard 2002; Buckley 2012), and architectural styles (Jordan and O'Neill 2010), or nonmaterial cultural products such as religious beliefs (Watts et al. 2015), kinship systems (Jordan 2013), music (Windram et al. 2014), and stories (Da Silva and Tehrani 2016). In these cases, the phylogenies directly represent a model of the way those cultural features evolve and diversify over time. Secondly, phylogenies can track the populations in which

✉ Lindell Bromham
Lindell.Bromham@anu.edu.au

¹ Macroevolution & Macroecology Group, Research School of Biology, Australian National University, Canberra ACT, Australia

culture evolves, for example by generating phylogenies from DNA sampled from people, pathogens, companion animals, or domesticated plants. These phylogenies are not directly based on the outcomes of cultural evolution, but are used to interpret cultural change and diversification, even in the absence of material evidence from the relevant places or time periods (e.g., Gray et al. 2011; Grollemund et al. 2015). Language phylogenies can fall into either category—language can be used to track cultural evolution, for example through kinship terms (e.g., Jordan 2011), or it can be used to track population history and movement (e.g., Gray et al. 2009).

This article is focussed entirely on the second class of phylogenies of human culture, which represent the evolutionary history of the humans themselves. This is not a review of phylogenetic methods, nor a summary of the evidence for or against particular hypotheses about human history. It is also not a review of the field of cultural phylogenetics, for which there are already many excellent overviews (such as Mace and Holden 2005; Gray et al. 2007; Mace and Jordan 2011; Mesoudi 2016; Evans et al. 2021). Instead, the focus is on a much broader question: how should we interpret phylogenies of human populations? Can they be read as literal histories? How can phylogenies be used to understand the processes of human movement and diversification? Here we consider the nature of phylogenetic inference, not the construction of phylogenies themselves.

Phylogenies and the Peopling of Australia

I have chosen three case studies that each explores the same topic—the peopling of Australia—but uses a different dataset to trace this history: human genes, virus genes, and words. In each case, a different “marker” is being used to track human history, because humans carry their genes, pathogens, and languages with them as they move over the landscape and diversify. Each one of these phylogenies is a remarkable technical achievement in its own right, and each opens a window on the evolutionary past and processes. This article is not a technical examination of these studies, nor a critique of their findings, nor is it a review of phylogenetic studies of Australian history. Instead, these three examples are chosen because they illustrate issues that are common to all such studies, including the application of simple models to complex histories, the challenges of reporting uncertainty, and derivation of a coherent historical narrative from a bifurcating hierarchy. We could have focussed on any similar collection of phylogenies and made exactly the same points, but these three are useful illustrations because of the diversity of data sources (human genes, virus genes, languages), the challenging nature of the question (deep history of Australian people), and the differences

of approach in interpreting and reporting the conclusions (although the methods used are similar). In pointing out the similarities and differences between these case studies, the aim is to highlight challenges faced by all phylogenetic studies of human history.

The first example is a molecular phylogeny of Indigenous Australians, based on mitochondrial DNA extracted from museum samples of human hair collected by anthropologists in the 1920s and 1930s (Tobler et al. 2017). Molecular phylogenetic analysis of mitochondrial DNA from these hair samples was used to estimate the history of movement of people across the continent, from their entry point on the northern coast to their occupation of areas in southern Australia (Fig. 1). Using rates of mutation calculated from other human populations, the authors obtained dates for human arrival in northern Australia between 47 and 43 thousand years ago, and suggested people dispersed and occupied sites in northeastern and southern Australia by within a few thousand years of arrival on the continent. These results were used to support a narrative of rapid migration following human settlement, with people migrating thousands of kilometers across a broad variety of landscapes and habitats.

The second phylogeny is based on the genomes of hepatitis B viruses (HBV) sampled from Indigenous Australians

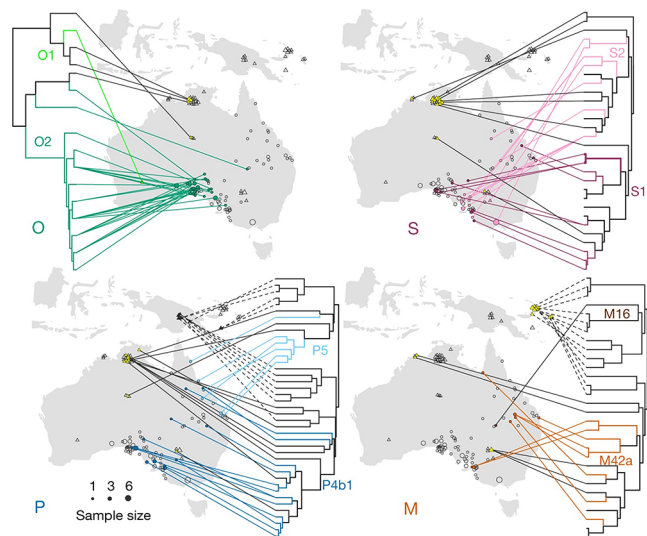


Fig. 1 Phylogenies of mitochondrial haplotypes (groups of distinct sequence variants labelled O, S, P, M) from Indigenous Australians, including both sequences extracted from museum hair samples and contemporary samples. Black dots represent the inferred location of the oldest known maternal ancestor of the person who provided the hair sample, based on information they provided about pedigrees and homelands. Triangles are modern samples reported from previous studies. (Reprinted by permission from Springer Nature Customer Service Centre GmbH: Nature, Aboriginal mitogenomes reveal 50,000 years of regionalism in Australia. Tobler R, Rohrlach A, Soubrier J, Bover P, Llamas B, Tuke J, Bean N, Abdullah-Highfold A, Agius S, O'Donoghue A, 2017)

analytical techniques that allow complex evolutionary models, and construction of comparative linguistic databases for hundreds of languages. Each of these studies also represents interdisciplinary collaborations, with authors including linguists, computational biologists, health professionals, virologists, anthropologists, statisticians, earth scientists, museum curators, and Indigenous community representatives.

If we are to take these three phylogenies at face value, reading each phylogeny as a temporal and spatial map of history, detailing a series of events where ancestral populations changed and diverged to produce an array of descendants, then combined they are consistent with a picture of humans having the intellectual and technical ability to cross the sea from Southeast Asia to Australia over 51,000 years ago, the adaptability and ingenuity to be able to rapidly spread over a large continent and occupy distinctly different, and often changing, ecological environments, and the flexibility that allowed a more recent wave of cultural change to sweep through the majority of the continent completely replacing older languages yet not displacing the resident populations. Yet the phylogenies also differ in their implied histories, suggesting different dates of origin of the Australian people, and different patterns of expansion. The purpose of this discussion is not to critique the data, methods, or results of these studies, but to use these three studies as exemplars in order to examine issues common to all such analyses. How do we translate these phylogenies, each based on groundbreaking data and sophisticated methods, into narratives about the past and processes of human diversification?

Phylogenies are an Abstract Representation of History

The first challenge in interpreting phylogenies as human histories concerns the most fundamental feature of a phylogeny, which is that the history of the lineages is represented by a bifurcating tree diagram. Representation of human history, or cultural and linguistic evolution, in terms of a branching tree has long been controversial, but is defended as a valid simplification of processes of diversification. We can use each of these case studies, reconstructing the peopling of Australia through words, genes, and viruses, to explore the way that bifurcating trees are a simplified representation of complex evolutionary processes.

The phylogeny of the Pama-Nyungan languages is a phenomenal achievement for understanding the evolution of Australian languages. As for all such phylogenies, the representation as a tree must involve abstraction from a more complex, intertwined history of change and exchange of words. To highlight the challenges of interpreting language phylogenies in terms of human history, we can focus on one

small part of this phylogeny, the Ngumpin-Yapa subgroup of languages from northwestern Australia (Fig. 3). Like all other parts of the phylogeny, the relationships between these languages are organized in a bifurcating hierarchy of nested two-way splits. But if we were to interpret these splits as exactly mapping to the history of the people who speak these languages, we would paint a strange and unfamiliar view of the way that human populations diversify. A literal interpretation would suggest that once upon a time (around 3000 years ago), there was a group of people that spoke Proto-Ngumpin-Yapa, then these people split into two groups. One of these groups spoke a Ngumpin language, and the other spoke Yapa. The people who spoke Ngumpin split into two again, one of which spoke Jaru Ngardi, and the other spoke a Gurindjic language. By and by, the Jaru Ngardi speakers also split into two lineages, one of which spoke Jaru and the other spoke Ngardily. And after some time, some of the Gurindjic speakers split into a group that became the Malngin speakers, then another group left and they became the Ngarinyman speakers, leaving the Gurindji to talk amongst themselves. The remaining folk split again into two groups, one group speaking Bilinarra and the other group speaking Mudburra. When we say it like that, reading from the series of nested two-way splits represented on the phylogeny, it sounds like an origin story. But there are several reasons that we would hesitate to read a language phylogeny as a simple story about people undergoing repeated rounds of two-way splits. Firstly, we don't expect humans to behave in such a neat way, always splitting into two, never splitting into three or more, never merging or rejoining, so this series of bifurcations must be a simplified, not literal, representation of the past. Secondly, even if populations of people were to follow a neat series of two-way splits, we know that words don't always follow the same lines. Words, like any aspect of culture, can be both inherited from a common ancestor and passed between contemporary populations. "Borrowing" of words has long been cited as a problem for language phylogenies.

The Ngumpin-Yapa languages illustrate the challenges posed by the complex histories of words within languages. Some of the languages in this group form a gradient of dialects, without clear boundaries between related languages (Meakins et al. [in press](#)), and some show a high degree of borrowing. Forty-nine percent of words in the Gurindji lexicon are considered to represent borrowings from neighboring languages (McConvell 2009). Neighboring languages Mudburra (Ngumpin-Yapa, PNY) and Jingulu (Mirndi, non-PNY) are an even more extreme example of sideways movement of lexicon, sharing 65% of their nouns and nearly 40% of their verbs with words borrowed in both directions (Pensalfini and Meakins 2019; Meakins et al. 2020). So the evolution of these languages has not been by a strictly

bifurcating process of diversification, but a more complex set of interactions as languages diverge, merge, move, exchange (McConvell and Bowern 2011). This complex history does not negate the phylogeny, but it serves to focus attention on the meaning we assign to the tree.

We can't capture the full complexity of these relationships in a single nested series of bifurcating splits. For example, the relationship between Jingulu and Mudburra can't be included in this phylogeny of Pama-Nyungan languages (Fig. 3), because Jingulu is from a different language family. Yet Mudburra shares more lexicon in common with Jingulu than many of its closer relatives. There is no simple way to represent this in a tree—on the one hand, the last common ancestor of Jingulu and Mudburra was over 6000 years ago, deep in the unknown history of these languages. On the other hand, particular words have a more recent shared history between Jingulu and Mudburra, as words have been exchanged through contact. This phylogeny is based on a sample of words from basic vocabulary, which are considered to be generally resistant to borrowing, so these words may be a better indicator of the deep history of language inheritance than the more recent history of borrowing and exchange.

Any phylogeny must be a simple representation of a complex history because it tracks only a handful of possible “markers” of history, and presents some kind of summary over the markers sampled. A language phylogeny is necessarily a depiction of the diversification of sample of vocabulary, and not necessarily representative of the whole lexicon. In microbiology, this has been referred to as “the tree of 1%,” a sample that reflects part but not all of the history (Dagan and Martin 2006). Other words may tell a different story; for example, the high degree of borrowing in Gurindji might be a reflection of a history of population movement and expansion (Bowern et al. 2011). The tree is a summary of patterns of similarity in a sample of words, but it is made on the assumption that this summary will provide a fair history of the languages and the people who speak them.

Nyumpin-Yapa languages might seem unusually evolutionarily permissive, but there are many language groups where a nested set of bifurcating splits is not a realistic description of the evolutionary history of the people who speak the languages. The first formal evolutionary trees of languages were for the Indo-European family (Schleicher 1863). When placed in a phylogeny of Indo-European languages, English is usually grouped with the Germanic languages, despite bearing the marks of substantial input from Norse, Celtic, and Romance languages. English can, on the basis of much of the core lexicon, be grouped with Dutch, Frisian, and German in a phylogenetic tree, but this doesn't mean that the history of the English language or the English people has occurred through a series of two-way

splits (Heggarty 2014). The history of English may be better represented as a network of some kind, showing multiple inputs from different languages, rather than a simple ancestor-descendent relationship within the Germanic clade (Nelson-Sathi et al. 2011). Words are carried along with people as they move, settle, and adapt, but words can also be gained and lost, borrowed and shared with other people who have different histories, creating a complex mosaic of histories tracing many different paths through invention, contact, change, and loss. For this reason, there has been a long history of skepticism of the depiction of language histories as bifurcating tree diagrams (see discussions in Kalyan and François 2018; Jacques and List 2019).

Some language histories may conform more closely to a phylogeny-like structure of a nested series of splits. A plausible example is the Austronesian languages of far Oceania, which includes the descendants of groups of people who split from their ancestral language groups and journeyed across the ocean, establishing new societies on previously uninhabited islands (Gray et al. 2009; Bromham et al. 2015). Even so, Pacific people continued to travel back and forth between islands, reconnecting the divergent lineages, and leaving an interconnected web of linguistic borrowing and genetic admixture (Ross 1988). Words have been borrowed not only between related Austronesian languages, but across greater linguistic divides, such as between Papuan languages and Austronesian languages, where they are in contact. A study of one group of Austronesian languages suggested that on average around 8% of words in basic vocabulary had been borrowed from a non-Austronesian neighbor (Robinson 2015), a figure broadly consistent with estimates for borrowing in basic vocabulary in many other language groups (Bowern et al. 2011; Greenhill and Gray 2012). Given that basic vocabulary (e.g., Swadesh lists) are considered fairly resistant to borrowing, this is likely to represent a lower limit of the borrowing rate. The frequency of borrowing, unsurprisingly, varies widely across different languages, but is in most cases not trivial, and many cases have reported borrowing levels above 10% of vocabulary (Tadmor 2009). Where there is borrowing between languages that are not each other's closest relatives, words in the vocabulary do not all share the same history: there is no single set of ancestor-descendant relationships that accurately represents the history of all words in the language. Borrowing does not invalidate phylogenetic analysis, though it can sometimes result in odd placements in a language phylogeny that do not reflect relationships based on other forms of evidence (Greenhill et al. 2010). But, if there is a non-trivial degree of borrowing, the phylogeny then must represent some kind of summary statement, reflecting both inheritance and sharing, that is not literally true for each element of the language.

There is a tendency to think of borrowing as a special problem of language evolution, a complication that biologists don't have to worry about. But this is far from the case. Evidence has been growing for genetic “borrowings” in the form of horizontal gene transfer—the movement of genes from one species to another, rather than inherited from a common ancestor of both. Like linguistic borrowings, horizontally transferred genes complicate phylogenies because not all genes in the genome can be represented by a single history encapsulated in a simple bifurcating tree (Doolittle 1999; Dagan and Martin 2006). Sometimes, like the transfer of words between Gurindji and Mudburra, genetic transfers are between close relatives. For example, the evolutionary history of the North African date palm (*Phoenix dactylifera*, domesticated over 7000 years ago), can't be represented as a simple branching diagram, because it contains genes that link it not only to its closest relatives but also to wild varieties from the eastern Mediterranean, presumably due to gene flow or hybridization (Flowers et al. 2019). Sometimes, like word swapping between Mudburra and Jingulu, genes are transferred between distantly related lineages. For example, most animals cannot make their own carotenoids and must therefore obtain them from a healthy diet (which is one good reason why you should eat your vegetables). But some pea aphids are a chirpy orangey-red color, like a well-known politician, because they can make carotenoids, a trick they picked up by borrowing the gene for carotenoid biosynthesis enzyme from a fungus (Moran and Jarvik 2010). If you make a molecular phylogeny of carotenoid biosynthesis genes, the aphid comes out nested within the fungi: the phylogeny identifies the history of the genes (most closely related to fungi), not the evolutionary origins of the aphids (most closely related to other animals). Different parts of the genome have different histories, just as different words in the lexicon do.

Genetic evolution is likely to look less tree-like the closer you look, particularly when considering relationships between populations within a single species, like humans. Just as the Pama-Nyungan tree describes the pattern of similarity of words in the lexicon, which do not all share the same history, genetic phylogenies describe patterns of similarity of genes in the genome, which do not all share the same history. Ancestral polymorphism and gene flow between populations create non-tree-like patterns, and can influence the relationships, timescale, and estimates of past population sizes (Leaché et al. 2013; Tiley et al. 2020). Tobler et al. (2017) attribute the intermixing of Melanesian and Australian sequences in the mitochondrial phylogeny as signs of incomplete lineage sorting—that is, that multiple genetic variants were present in the ancestral population, and that some of these have persisted in both Papuan and Australian populations to the present day. This means that,

like words that trace different lexical origins, the gene variants present in the modern populations don't all trace the same history, so reconstructing a tree for different variants would imply different times, places, origins, and relationships. Introgression—the movement of genes between populations—could, like borrowing of words, generate patterns of similarity that cannot all be fitted to a single bifurcating tree. Australian people have not been isolated throughout their history (Rowland 2018). For example, the dingo came to Australia from New Guinea within the last ten thousand years (Balme et al. 2018), and its introduction is typically interpreted as human-mediated via seagoing travelers from Indonesia or New Guinea (Fillios and Taçon 2016). So it is possible that markers of human history in Australia, including human genes, viruses, and words, have a tangled history, rather than a simple series of bifurcating splits.

Hepatitis genomes were chosen to trace the history of Australian people because they change slowly, and are carried along with human populations, just like human genes are. But viruses also can have complex patterns of evolutionary history, including not only splitting of one lineage into two, but merging of two genomes into one. We can see an example of merging lineages in the Australian hepatitis genomes, because one of the viral strains included in the phylogeny that was used to date the peopling of Australia is a recombinant virus, a hybrid formed by joining parts of the genomes of two different virus strains (Yuen et al. 2019). A recombinant genome, like borrowed words or horizontally transferred genes, cannot be described with a single evolutionary history, because different parts of the genome trace different evolutionary histories. In the phylogeny based on the whole genome, the hybrid strain clusters with the lineage that donated the larger portion of the genome, grouping the HBV/C4 virus with the Asian HBV/C lineages (Yuen et al. 2019). But if you infer the phylogeny from the surface genes only, the relationships are entirely different: HBV/C4 groups with HBV/J (previously isolated from a Japanese patient) and the nonhuman-primate viruses from Southeast Asia (Fig. 2). There are two distinctly different histories written in the same viral genome, which cannot be simply represented on a single bifurcating tree.

Borrowing, contact, horizontal transfer, hybridization, introgression, or admixture result in complex evolutionary histories that cannot be represented by a single branching diagram for genes, pathogens, and words. While there has been much debate about the degree to which culture or language will show tree-like structure, the same is also true for genetic data, so it has been argued that the fit of cultural data to a phylogenetic pattern may be no worse than for many biological datasets (Collard et al. 2006; Gray and Watts 2017). A phylogeny can at best be an abstract representation of the process of evolution, whether cultural or

genetic. It is important to emphasize that the simplicity of trees compared to the complexity of evolution is not a fault, and does not mean that the application of phylogenies to genes, languages, or pathogens is doomed to failure. All of these trees are informative and useful. As we will see in the second half of the article, for many applications, an abstract representation is adequate, and any information on evolutionary relationships is usually much better than none at all. Even if we think that a bifurcating tree is not a wholly accurate description of the history and relationships of a set of cultures, it will usually be a *better* representation than assuming there are no patterns of similarity due to relationships at all (Levinson and Gray 2012; Paradis 2014).

Phylogenies Reflect Assumptions About History and Mechanisms of Change

We have seen that a phylogeny should generally be regarded as a simple abstraction of a complex history. This is not a flaw, but a necessary part of using phylogenetic methods to represent the evolutionary past and processes for a set of lineages, using a sample of relevant data. This abstraction is common to all such studies, independent of the data used or the methods employed. The reliance on a set of simplified assumptions about the evolutionary process (such as diversification occurring by bifurcating splits) is an unavoidable part of phylogenetic inference, whether or not those assumptions are stated explicitly, and many of these core simplifications are independent of the particular statistical framework used to generate the phylogenies. The sophistication of phylogenetic methods has changed hand-in-hand with the size of datasets and computational capacity. One of the outcomes of the ever-increasing sophistication of phylogenetic methods is that it is now very difficult to succinctly express all of the assumptions underlying the analysis (Bromham 2019). The complexity and flexibility of the methods, and the wide range of parameters used, also means that a wide range of outcomes can be obtained by changing the assumptions (Bromham et al. 2018a). Therefore one of the things we must bear in mind as we interpret phylogenies as human histories is the degree to which the answer we get is shaped by the assumptions we make.

One of the important assumptions underlying most genetic phylogenies, and an increasing number of linguistic and cultural phylogenies, is that we can infer rates of change over time. It is only by assuming or estimating absolute rates of change that we can read dates of origins of different lineages. For their analysis of mitochondrial genes, Tobler et al. (2017) assumed a prior mutation rate estimated for other human populations. For the virus and language phylogenies, rates were estimated by assuming the age of one particular split in the tree, and using that to estimate rates of

change for the rest of the phylogeny. Both approaches rely on having an accurate model of the way rates behave in different lineages, whether it remains constant or changes over the tree (Welch and Bromham 2005). Both approaches also rely on the accuracy of any known date of divergence used to calibrate rates of change, whether for an assumed prior estimate or estimates generated in the phylogenetic analysis (Gray et al. 2011).

Consider the hepatitis B phylogeny that was used to infer the movement of people into Australia over fifty thousand years ago (Yuen et al. 2019). A phylogenetic tree was inferred for 59 whole genome sequences of HBV from Aboriginal people from northern and Central Australia and Melanesian people from Vanuatu, along with other human and nonhuman-primate HBV sequences. To infer the timing of the movement of HBV into Australia, a rate of molecular evolution was calculated by assuming the age of the Asian and Melanesian lineages corresponded to earliest evidence of anatomically modern humans in each region (34.5 and 40 thousand years ago, respectively), on the assumption that HBV sequences tracked the movements of humans out of Africa and as they spread around the world. As would be expected, estimates of the rate of evolution of HBV varied across the different lineages within this study. More broadly, the rate estimates also differ between different published studies using HBV to track human movement. In a different study of HBV evolution and diversification, archaeological evidence of human migration into the Americas was used to estimate a much faster rate of change for the hepatitis genome (Paraskevis et al. 2013). Using this faster rate results in an inferred date of origin of the Australian strain of less than seven thousand years ago.

Regardless of which dates are considered more plausible in light of other forms of evidence, these contrasting dates illustrate how essentially the same data can be used to support different conclusions depending on the assumptions of the analysis (in this case by using different calibrations). Because of this, it is often the case that different published studies produce quite different timescales for the same history. For example, two molecular dating studies of Australian people published in the same journal only a year apart came up with quite different date estimates: while Tobler et al. (2017) suggested a colonization and expansion starting from around 50 thousand years ago, Malaspinas et al. (2016) proposed a much younger date for the expansion of people across Australia, suggesting Pama-Nyungan-speaking Australians are all descended from a population expansion between 10 and 35 thousand years ago (Malaspinas et al. 2016). These studies differ in data used, but also in many of the assumptions underlying the analysis. A wide range of assumptions are made in any phylogenetic analysis, concerning rates of change, calibrations (known

dates), and evolutionary mechanisms, and these can all influence outcomes of phylogenetic analyses (e.g., Ritchie and Ho 2019). For example, employing different models of lexical change to analyze Sino-Tibetan language data (comparing a model of cognate gain and loss to a covarion model) resulted in a 9000 year difference in inferred age of origin, and changing the model from a single rate of change to a multiple-rate “relaxed clock” changed the date by over 1500 years (though the confidence intervals on the dates overlap) (Zhang et al. 2020). The dates derived depend on the assumptions made, but often the precise nature of the assumptions is made opaque by the sophistication of the methods (Bromham 2019).

Decisions made in the analysis, such as calibrating information or models of evolutionary change, can have a substantial influence on the results. Some of these assumptions can be debated in terms of empirical evidence, such as the age of particular nodes in the tree, or in terms of the reasonableness of alternative choices, such as which nodes to use as calibrations (Paraskevis et al. 2013; Yuen et al. 2019). Others may be subject to ongoing revision as new research emerges, such as the human mutation rate (Tobler et al. 2015; Malaspinas et al. 2016). But in the case of some core assumptions of phylogenetic analyses, choices may be predominantly based on investigator preference with little empirical grounds on which to choose one assumption over the other, such as clock models or “tree priors” (that describe the mode of diversification) (Bromham et al. 2018a; Ritchie and Ho 2019). Choice of models and formal statements of prior assumptions are just some of the ways in which investigator choices influence the outcome of phylogenetic studies.

Phylogenies are Subject to Interpretation

Phylogenetic analysis has a satisfyingly objective feel to it: gather the data, run the analysis using the best possible methods available, report the results as statistical estimates with confidence intervals. But subjectivity—dependence of conclusions on the researchers conducting the analysis—comes into all phylogenetic analyses in three important ways: choice of data, methods, and assumptions; identification of the optimal phylogenetic solution; and interpretation of that solution in terms of evolutionary history and processes. Subjectivity is not a bad thing—this is where expertise comes into play—but it is important that it is recognized when interpreting the results of phylogenetic studies.

We have seen one way that researcher subjectivity influences the results of phylogenetic analyses: choices made in analysis, such as calibration, can influence the inference of history from phylogenies, and different assumptions can produce quite different answers. For example, calibrating

the rate of change of the HBV genome using evidence of modern human arrival in Asia gives a date of colonization of Australia over 51,000 years ago (suggesting early establishment without later replacement), but calibrating the analysis on human arrival in North America gives a date of origin of the Australian HBV strains less than seven thousand years old (suggesting an influx of viruses in the more recent past, long after human occupation of the continent). Researchers make other important choices in phylogenetic analyses: choosing which solution to report, and interpreting the meaning of that solution. These choices, which are an important part of the expertise that researchers bring to a phylogenetic project, can generate considerable leeway in the degree of support that a study gives to a particular hypothesis about human history or cultural evolutionary patterns.

Even a single analysis with one set of assumptions typically produces a large range of equally plausible phylogenetic solutions. Deciding how to convey the outcomes of the analysis given this range of plausible solutions is one of the choices that researchers have to make. One option is to present all of the equally plausible solutions. For example, the HBV phylogeny is plotted as a “densitree” that shows a cloud of possible solutions which show a range of topologies (solutions differ in the relationships between strains) and depths (solutions differ in the ages of lineages). In this way, the set of plausible dates of the base of the Australian HBV sequences can be seen to vary over tens of thousands of years (Fig. 2). A similar figure is given in Tobler et al. (2017), where the range of dates covering the earliest divergence events covers tens of thousands of years. An alternative approach is to produce a summary statement of the set of plausible solutions. The Pama-Nyungan analysis also produced a large sample of possible trees, but, in common with most phylogenetic studies, for simplicity they present a single phylogenetic tree to indicate the solution with the highest support, often referred to as a consensus tree (they use a maximum clade credibility tree, which is a way of summarizing the most common groupings from all the trees in the sample). The relationship between a consensus tree and the sample of trees on which it is based is illustrated in Fig. 2. Support values on this summary phylogeny can indicate how many of the equally plausible solutions contained the same groups, such that some of the groupings are found in all trees in the sample, some in fewer. For example, relationships within the Ngumpin-Yapa group vary in support from 0.31 (supported in a third of all sampled equivalent phylogenetic solutions) to 1.0 (in all sampled trees: Fig. 3).

For some phylogenetic studies, decisions are made to exclude some possible phylogenetic solutions as implausible. For example, in the study of Australian history using mitochondrial DNA, the groupings by haplotype were

constrained, meaning that any solutions that did not support separation of these defined groups were not considered in this analysis. This study presents a single phylogeny but the distribution of possible ages for each haplotype group is shown, based on the constraints on haplotype groupings. Support values for the nodes in this tree are provided in extended data accompanying the paper, and show a range of values, from 1 (present in all of the most plausible solutions) to 0.04 (present in 4% of the solutions).

Decisions about which assumptions to make in the analysis, which solutions will be considered valid, and which tree to present as the final results help to shape the conclusions drawn from the study. But the story told is also strongly shaped by the way that the authors read the phylogeny. These papers would be a lot less interesting to read if they simply presented a phylogeny without further interpretation. In each of these three studies, the authors present a more easily digested graphical summary of their conclusions in the form of a map with arrows of suggested paths of human movement (Fig. 4). These diagrams are informed by, but not wholly contained in, the phylogenies. This is not a criticism: the phylogenies provide useful information but they do not provide a coherent historical narrative, so it is up to the researchers to interpret the phylogenetic information in terms of what they consider a plausible explanation. To draw a slightly tenuous analogy, if you want to draw a dinosaur, you have to give it a color and skin texture, even if all you have are bones: you can't draw the whole animal without inferring the parts you can't observe.

Examination of the way the historical narrative is constructed in these three papers, based on but not wholly

contained in the phylogeny, is not intended as a criticism of these studies, but as an illustration of a source of subjectivity inherent in all similar phylogenetic analyses: how to get from the tree to the history. For example, the diagram of coastal migration down the west and east coasts presented by Tobler et al. (Fig. 4a) is only partly informed by the mitochondrial DNA analysis, which includes only one sample for the western half of the continent (the hair samples were taken from three locations, and the map positions of the samples are based on self-reported information on each sampled individual's personal or family history). The movement from north to south is a plausible inference given that people are most likely to have entered Australia from the north coast, but it does not correspond directly to information contained in the phylogenies, which by themselves provide no more evidence for north-south movement than for south-north or east-west (Fig. 1), nor is it directly attested in the correlation tests applied to the data, which provide various levels of support to patterns of variation along both latitude and longitude.

Similarly, two paths of movement of HBV from the northern coast are shown on the grounds that they form two distinct clades. While this is a reasonable explanation of the patterns in the data, the phylogeny itself does not provide a clear pattern of migration and many of the clades are incompletely resolved (for example, the separation between the C4a and C4b groups was only clearly supported in 62% of solutions), making the exact patterns of diversification ambiguous given these data. Other interpretations, such as a movement from western Queensland to central Australia to the northern coast, would also be possible given this phylogeny. So the reasonableness of the solution presented in Fig. 4(b) depends not wholly on the phylogeny but also on the plausibility of alternative histories based on additional information. I emphasize that drawing attention to subjectivity is not a criticism of the conclusions, which may well be the most reasonable explanation of the observed data. But it serves to illustrate that the historical narrative is drawn not wholly from the tree but also from other sources. This is important to recognize because the veracity of the history cannot be evaluated simply by measuring the strength of support for the tree, without recognizing that the story also depends on additional non-phylogenetic information.

The Pama-Nyungan language paper also provides a graphical summary of the results by plotting the phylogeny on a map, using analytically inferred ancestral locations for placement of nodes (Fig. 4c). This image could be read as tracing paths of movement across the continent. For example, tracing the yellow line would suggest that node *o* (common ancestor of the Arandic languages) gave rise to a lineage that moved west to establish *h* (common ancestor of the Wati languages), and that this lineage gave rise to the

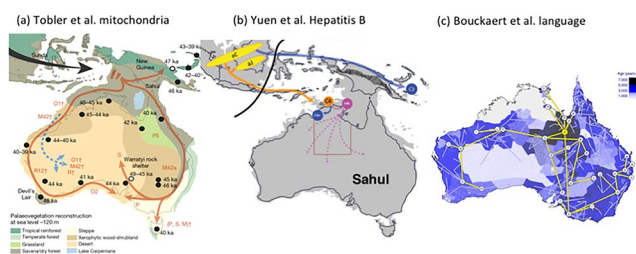


Fig. 4 Graphical interpretations of results of phylogenetic studies from (a) Tobler et al. 2017 (see Fig. 1); (b) Yuen et al. 2019 (see Fig. 2); (c) Bouckaert et al. 2018 (see Fig. 3). (Reprinted by permission from Oxford University Press (Yuen LKW, Littlejohn M, Duchêne S, Edwards R, Bukulatjpi S, Binks P, Jackson K, Davies J, Davis JS, Tong SYC, Locamini S, Tracing Ancient Human Migrations into Sahul Using Hepatitis B Virus Genomes, *Molecular Biology and Evolution*, 2019, Volume 36, Issue 5, pages 942–954) and Springer Nature Customer Service Centre GmbH (Nature, Aboriginal mitogenomes reveal 50,000 years of regionalism in Australia. Tobler R, Rohrlach A, Soubrier J, Bover P, Llamas B, Tuke J, Bean N, Abdullah-Highfold A, Agius S, O'Donoghue A, 2017; *Nature Ecology & Evolution*, The origin and expansion of Pama–Nyungan languages across Australia, Bouckaert RR, Bowern C, Atkinson QD, 2018))

k node (Ngayarta languages), which in turn gave rise to n (Kartu) then m (southwest languages). However, while this may well be a fair representation of the history of Australian diversification, it is not directly attested in the phylogeny itself, in which all of these nodes are of equivalent rank, rather than nested within each other. In the tree, n , m , and h are each other's relatives, lineages that all emerged from a common ancestor. Along with i , j , k , and l , it is not possible from the phylogeny alone to say that any one of these groups gave rise to any of the others.

The aim of this discussion is not to cast doubt on the conclusions of these three studies, but simply to point out that the conclusions of phylogenetic studies are not simply read, objectively, from the phylogeny itself. The data does not have a voice of its own, it must speak through researchers' ideas, decisions, and beliefs. The narrative told from phylogenies depends on the data collected, the assumptions made in the analysis, and the researchers' interpretation of the results. The phylogeny provides the scaffold upon which the researchers build the story, which may follow the lines of the scaffold more or less rigidly. The veracity of the conclusion rests not just on the phylogeny but all the auxiliary beliefs that shape the narrative, and those beliefs are based on expert opinion. Where expert opinion differs between researchers, then the narratives based on the phylogeny may differ as well. Researchers interpret phylogenies in light of what they believe to be a reasonable explanation for the observations they have made. The story told is informed by, but not wholly inherent in, the tree.

Phylogenies are Stories

Even though the results of phylogenetic studies are often described as “estimates,” they differ in important ways from most statistical estimates (Bromham 2019). An estimate is usually considered to be a value derived from a sample of observations in order to make a statement about the likely value for the population as a whole. For example, we could estimate the levels of genetic variation in a population by sampling many individuals and using the observed levels for genetic variation to generate an average value for the whole population. The more individuals we sample, the more we expect our estimated value to approach the true value (i.e., the actual level of genetic variation, if we knew the status of every individual without error). But when we use that genetic data to make a phylogeny, we are doing something rather different: we are using the patterns observed in the data to suggest past events that are not accessible to direct observation. None of the data—whether gene sequences, words, or cultural objects—are direct observations of the target of our investigation, which is the past events that produced our data. Instead, we rely on making assumptions

about the processes that produced these observations in the present day to allow us to reconstruct the past from our data. However simple or complex, explicit or implicit, assumptions about evolutionary processes are an essential part of historical inference.

This is abductive reasoning: using what we understand about the way the world works to come up with a plausible explanation for the observations we have made. It is principled storytelling. We can never know for sure whether we have correctly reconstructed an unobservable past event; the best we can do is use all of the information we have to make the most plausible explanation we can, and to subject that explanation to testing against alternative explanations or new information. This property of phylogenetic analysis is shared with all kinds of historical inference. For example, Currie (2018) uses the example of an extinct platypus to explore the way that historical scientists use evidence to reconstruct the past. *Obdurodon tharalkooschild* is known only from a single tooth, millions of years old, but features of the tooth were sufficient to identify it as a relative of the modern platypus, only much bigger. Differences between this tooth and modern platypuses were considered to indicate that it took larger prey items such as frogs and turtles (Pian et al. 2013). This picture of a giant, carnivorous platypus may well be true, but it is model-based historical inference rather than direct observation. The researchers used assumptions about the way the world works (such as the relationship between tooth size and body size and the association between dental features and diet) to infer features of an extinct organism that cannot be directly observed (such as body size and foraging behavior). Currie (2018) claims that these features effectively allow the historical scientists to generate additional evidence. But, while we can use these reconstructions to evaluate hypotheses about the past, the nature of this evidence is of a different kind from the factual observations made in the present day. The tooth is an undeniable fact—we can hold it in our hands. The giant platypus is the story we tell to explain the tooth. It may well be a fair description of a once-living animal, but we don't know for sure because we can't directly observe the animal itself (Currie 2018). In the same way, the phylogeny is the story we tell to explain the genetic data, or language data, or virus genomes that we have observed. It may well be true, but without additional information we cannot know for sure (Bromham 2019).

Descriptions of results of phylogenetic analyses as historical inference (Yuen et al. 2019) or in terms of support for hypotheses (Bouckaert et al. 2018) are better reflections of the role that phylogenies play in forming and evaluating ideas about the evolutionary past and processes. There is a practical reason for making this distinction between estimates and inference. Estimates get more accurate and more

precise with increasing amounts of data, but this is not necessarily the case for inference. We can sequence the whole genome of every individual in the population, or record every word in the language, and we will still get the wrong answer about the nature and timing of past events if our assumptions about how the sequences or languages evolved are wrong. For example, adding more HBV genomes would not change the disagreement in dates generated by using different calibrations, because the disagreement between the studies comes down to which calibration dates to believe.

There is another reason that adding more data won't solve necessarily improve accuracy and precision in phylogenetic analysis. We have seen that for most kinds of evolved entities—including languages, cultures, and genomes—there is usually no single “true tree” that describes the history of all the data. As the amount of data increases, the phylogeny becomes a kind of average statement about the history, which may not exactly correspond to the history of any given element. In many cases, a nested series of two-way splits is an abstract representation of the processes of diversification that produced the outcomes we observe, and different elements within the data set may have different histories (such as borrowed words or introgressed genetic variants).

We have seen that phylogenies can rarely be read simply as an accurate and literal history. But phylogenies don't have to represent some kind of idealized “true tree” in order to play an important role in understanding human cultural evolution. Each of the phylogenies discussed in this article is impressive in the scale of the dataset and the sophistication of the analysis, each represents an exciting new way of gazing into the past and evaluating ideas about the tempo and mode of human history and cultural evolution, and each presents a plausible explanation of current diversity. In the remainder of the article, I will argue that phylogenies don't have to be “true histories”—factual representations of past events—in order to be extremely useful. More than just useful, phylogenies are indeed essential to many of the kinds of questions we want to ask about the evolution of human culture.

Purpose: Why are Phylogenies so Useful?

In the first half of the article, I have considered some of the challenges we face when we use phylogenies based on inheritable markers of ancestry—such as genes or languages—to understand the patterns and processes of human history. The points I have raised are well known and understood by people who produce and interpret phylogenies. So, given these limitations, why are phylogenies so ubiquitous? In the second half of the article, I will examine why phylogenies are so useful in studies of human history. The most

important point to make is that a tree does not have to be a literal and accurate depiction of history in order to be very useful.

If a language tree or molecular phylogeny is not a literal history of the people who speak the languages or carry the genes, then what is it? It is a description of patterns of similarity and difference in the data, and a model describing a possible way that those differences came to be. A phylogeny is a simplified explanation that captures just a few basic processes, rather than the whole complex history of variation and change. As a model, it does not need to be literally true in order to be useful. In fact, the explanatory power of a model is often due to its simplification of a complex reality.

Consider the way that population genetics—the mathematical treatment of the fate of gene variants in populations—revolutionized evolutionary biology. Mathematical population genetics was harshly criticized for relying on simplifying assumptions, and was dismissed by some as “beanbag genetics” (i.e., treating genetics as if it was as simple a process as drawing colored beans from a bag). The architects of the population genetic models were accused of ignoring the complexities of inheritance such as epistasis (genes influencing each other's expression patterns), linkage (genetic variance being inherited together, not sampled independently), and dominance (some genetic variance being expressed in preference to others). One of the most prominent evolutionary biologists at the time, Ernst Mayr, said, “To consider genes as independent units is meaningless from the physiological as well as the evolutionary viewpoint” (Mayr 1963). This is to confuse the model, which is equipment designed to do a particular job, with a belief about the nature of reality. The developers of population genetics were fully aware of these complications of real genetic systems, but they found they could make a lot of progress by working with a simplified and abstract notion of inheritance (Crow 2001; Dronamraju 2011; Rao and Nandjiah 2011). One of the architects of this mathematical framework, J. B. S. Haldane, explained that they “made simplifying assumptions which allowed us to pose problems soluble by the elementary mathematics at our disposal, and even then did not always fully solve the simple problems we set ourselves” (Haldane 1964). Treating heredity *as if* it was the same as drawing beans from a bag was a practical and effective approach to solving a complex problem (Hull 1988). Furthermore, it was a step in the right direction, not an endpoint in itself. Simplification may not always be necessary or desirable, but in some cases it can give you a large amount of explanatory bang for your buck (Matthewson 2011).

In the same way, evolutionary biologists can get a lot of explanatory value by treating evolutionary history *as if* it consists of a nested series of bifurcating splits. In both

population genetics and phylogenetics, abstraction to a simple model may be necessary due to incomplete knowledge, but it is also useful in terms of a workable solution that is sufficient for a wide range of purposes. Of course, there will be cases when the model is insufficient. For example, classic “beanbag” population genetic models may not be adequate to understand cases of “genetic anticipation” like Huntington’s disease, where the nature of the allele and its expression can change with each generation (McInnis 1996; Friedman 2011). We will also find cases where the simple tree model of a series of bifurcating splits will not be adequate, such as the origin of contact languages (Bryant et al. 2005; Daval-Markussen and Bakker 2011; Bakker et al. 2017). But for a lot of purposes, representing evolutionary past and processes *as if* they consisted of a series of bifurcating splits is a model that produces useful and informative outcomes. It does not have to be literally true to be useful, just as population genetic models that treated genes as independent “beans” were not literally true but were nonetheless very useful.

Haldane (1964) referred to the mathematical formulation of population genetics, with its simplifying assumptions, as a kind of scaffolding upon which evolutionary theory could be built up. Trees likewise provide a scaffolding for investigating evolution, allowing us to ask some questions that would be otherwise very difficult to investigate. In this section, I will outline some of the ways in which phylogenies can shine a light on the processes of human history and diversity. In each case, I will argue that even an imperfect phylogeny—one that abstracts complex processes to simple patterns, and may contain some errors or omissions—can be an amazingly useful and indispensable part of the researcher’s toolkit. The focus here is on the uses of phylogenies that trace human history through the incidental record left in genes or languages, rather than phylogenies of cultural traits themselves.

Phylogenies Give Shape to Processes of Diversification

The first advantage of phylogenies is one that is so obvious that it’s bound to be overlooked. Phylogenies provide a model for the generation of diversity. The fact that we can, by and large, recognize nested hierarchies of variation is evidence for Darwinian evolution (Romanes 1897). The only illustration in the *Origin of Species* is a branching diagram that depicts a simple model for the process whereby a single ancestral stock can give rise, over time, to many and varied descendants (Darwin 1859). The lines in Darwin’s diagram could be populations, species, orders, languages, technologies. It had been recognized long before any formal evolutionary theories were developed that species and

languages could be arranged in nested hierarchies (Pietsch 2012; Archibald 2014), but the Darwinian theory of evolution explained why this should be so:

All the foregoing rules and aids and difficulties in classification are explained, if I do not greatly deceive myself, on the view that the natural system is founded on descent with modification; that the characters which naturalists consider as showing true affinity between any two or more species, are those which have been inherited from a common parent, and, in so far, all true classification is genealogical; that community of descent is the hidden bond which naturalists have been unconsciously seeking. (Darwin 1859, p. 402)

A tree is not the only possible representation of the evolutionary process, and many alternative models for depicting diversification have been proposed. For example, the quirky quinary system arranged biodiversity as an interconnected ring of overlapping classes, which had some advantages in explaining organisms that appeared to be a bit like one class and a bit like another (Archibald 2014). Non-branching representations of evolution have a long history in linguistics, with the “wave model” published not long after the first branching trees (see François 2014). Ancestor-descendant lineages and reticulating networks have also played important roles in depicting cultural, linguistic, and biological evolution (Bryant et al. 2005; Geisler and List 2013; Morrison 2015). Tree-like depictions have two important advantages. Firstly, the tree model provides a simple model that solves for the central problem of diversity: why are there so many different species and languages rather than a continuum of variation or a linear process of anagenetic change over time? Not all diversity generation is by splitting; joining also can produce new varieties, such as hybrid species or contact languages. But phylogenies represent a process that results in the gradual accumulation of diversity over time. Secondly, in a practical sense, the tree model is extraordinarily useful as a simple and tractable way to describe similarity due to descent. No other representation has yet proved as useful as a branching, hierarchical arrangement. It doesn’t fit all cases (Creole languages and hybrid species being cases in point) but it is thus far the most successful general model that has allowed us to study the patterns and processes of diversification.

By representing the process of diversification, phylogenies allow comparison of rates and patterns of diversification over time, in a framework that can, at least in principle, be applied to all lineages, something that is not practical by any other means. Consider the question: do languages continue to diversify, producing ever increasing numbers of languages in the world, or is the addition of new languages

balanced by the loss of existing languages such that the number does not increase over time but either remains stable or decreases? To answer this, we need a way of estimating how many languages existed at different times in the past. Documentary evidence of long-ago languages is, like the fossil or archaeological records, patchy for many language groups and nonexistent for others. Because of the patchy and biased nature of historical record, we cannot estimate the number of past languages based on only on the fragments of direct evidence of ancient languages available in the present day. But a phylogeny allows inference of past diversity, even if it is not directly observable, because it makes the assumption that all the present-day languages evolved by a process of splitting from ancestral languages. The lines on a phylogeny connect the present to the past through hypothetical ancestral languages. To infer the number of languages that must have been present at a past time, we can draw a time slice across the phylogeny and count the number of ancestral languages that the phylogeny implies at that point in time, or plot the way the number of inferred lineages accumulates over time (Beyer et al. 2019; Hamilton and Walker 2019). Of course, this is subject to the same raft of assumptions as any phylogenetic analysis, concerning bifurcating splits, rates of change, and expected rates of both origination and loss of languages. And it cannot directly count past languages that have left no descendants or no documentary evidence in the present day. But it is one of the only general ways of estimating past language diversity, even if an imperfect one (Pagel 2009).

Dynamics of diversification can be inferred from phylogenies by asking whether nodes in the tree correspond to different time periods. For example, longer branches on the Austronesian language phylogenies were used to infer “pauses” in the colonization of Oceania, whereas series of shorter branches were taken as a sign of “pulses” of expansion, colonization, and diversification (Gray et al. 2009). Alternatively, the overall distribution of nodes throughout the phylogeny can be interpreted in terms of changes in diversification rate over time, or between different lineages. For example, the Austronesian and Pama-Nyungan trees were used along with eight other language phylogenies to track the accumulation of language diversity over time. A “lineages-through-time” plot was drawn for each, plotting the number of inferred lineages in the tree at each time slice, in order to ask whether the diversity accumulates steadily over time. On the basis of this plot, the authors suggested that the diversification of both Pama-Nyungan and Austronesian language families slowed over time (Hamilton and Walker 2019). Although some families increased in diversification rate over time, the overall picture was one of a slowdown in diversification rates as diversity increases.

Macroevolutionary analyses of diversification rates from phylogenies are subject to the same health warnings as any other phylogenetic analysis, regarding the potential uncertainties in relationships and dates. All of these analyses make strong assumptions about the way that rates of change vary over the phylogeny, about evolutionary processes such as extinction of lineages, and about the way contemporary diversity is sampled (Bromham et al. 2018). For example, phylogenetic analyses of macroevolutionary patterns rely on adequate characterization of contemporary diversity, which might be systematically biased in areas that suffered high degrees of language loss in recent times (such as Australia). Variation in rate of change across the phylogeny can lead to false inference of diversification rate dynamics (Shafir et al. 2020). In particular, if lineages with a faster rate of lexical change have a greater rate of producing new languages (Atkinson et al. 2008), then macroevolutionary methods might falsely infer a slowdown in diversification even when diversification rates are speeding up over time (and vice versa) (Duchêne et al. 2017). But, despite these limitations, there is nothing else that approaches the usefulness of phylogenetic analyses for tracing the history of diversification of the world’s languages and cultures (Jacques and List 2019; Evans et al. 2021).

Phylogenies Trace Paths of Gain and Loss

In addition to giving shape to the general pattern of diversification over time, phylogenies allow hypotheses about specific evolutionary processes to be put to the test. Consider an innovative application of the Pama-Nyungan language phylogeny to testing a long-standing hypothesis about the development of color terms in languages (Berlin and Kay 1969). This hypothesis has two core claims: first, that languages gain color terms by progressive addition in a predictable order (starting with black and white, then red, then yellow and green, then blue and brown, last purple, pink, gray, and orange); and second, that once gained, color terms are rarely lost. Directionality, repeatability, and reversibility of change are ideally suited to phylogenetic investigation. Haynie and Bowern (2016) compared the relative fit of different models of gain and loss of color terms along the Pama-Nyungan phylogeny, comparing a model with no losses to reversible models where terms can be gained or lost. They found that while gain was more common than loss, some loss of color terms was supported. They then compared the inferred rates of change between different color coding systems, finding broad support for the ordered gain of color terms, for example suggesting that red is typically (but not always) gained before green, and that brown and blue are only gained after yellow (Haynie and Bowern 2016).

Like any other phylogenetic analysis, this hypothesis test rests upon many assumptions. In addition to accurate coding of the presence or absence of color terms in each language (including the requirement that absences are true absences, not lack of information on relevant terms), this analysis relies on the tree being a map of history so that gains and losses can be correctly reconstructed. As we have seen, a bifurcating tree is an abstract representation of evolution, and may contain inaccuracies. Phylogenetic uncertainty casts doubt on results only if it can be shown that using an alternative tree would change the results. If all reasonable phylogenies support the same results, then we can be confident that the result is general. Haynie and Bower (2018) ran their analysis on a sample of 700 alternative trees that all represent equivalently plausible solutions (drawn from the posterior distribution of the original phylogenetic analysis). A more wide-ranging test of the robustness of the conclusions would try alternative phylogenies based on different assumptions, for example by changing the form of the “tree prior” (a simplistic model of speciation and extinction that defines the kinds of phylogenies considered most plausible solutions) (Bromham et al. 2018a; Ritchie and Ho 2019). Although it does not involve the underlying tree, an example of this kind of approach was applied following criticism of coding the presence or absence of color terms (Nash 2017): the data was reanalyzed with several alternative datasets and the conclusions upheld under all the various re-codings (Bower and Haynie 2017).

Phylogenetic analysis opens up modes of hypothesis testing that were previously unavailable. However, as with any model-based inference, the results only make sense in light of the assumptions made in the analysis. To illustrate the way that our conclusions could change depending on the assumptions we make, consider a phylogenetic analysis of a sample of 96 Austronesian cultures that was used to test the hypothesis that belief in “moralizing high gods” can enable evolution of greater political complexity through their role in suppressing selfish behavior (Watts et al. 2015). Of the six cultures that had moralizing high gods, three had high political complexity, and three did not. This looks indistinguishable from the outcome of a coin toss, and yet a phylogenetic analysis suggested a significant association between the two. To see why, we need to consider how the phylogeny is taken into account when evaluating the chance of two states co-occurring. High political complexity is more prevalent in this dataset than moralizing high gods (22 to 6) and both are highly labile, meaning they are highly changeable over the phylogeny, with multiple origins of each trait. A parsimonious interpretation of the phylogeny suggests over a dozen origins of high political complexity and half a dozen origins of moralizing high gods. These analyses rely on reconstructing the likely states of unobservable ancestral cultures, and

from there, infer how many times each trait has been gained in the presence of the other trait. Because high political complexity is gained so many times in the phylogeny (or, alternatively, gained fewer times and lost many more times), the ancestral state reconstruction inherent in the phylogenetic analysis infers that every branch in the phylogeny has a nonzero probability of being of high political complexity, even though this state is relatively rare across the whole tree. Since every ancestral lineage is inferred as possibly having high political complexity, the method suggests a significant relationship between moralizing high gods and political complexity even though there are only three actual incidences of co-occurrence. Similarly, broad supernatural punishment occurred in 36 of the cultures, and ten cultures had both broad supernatural punishment and high political complexity, which again is indistinguishable from a chance outcome of a random draw. Broad supernatural punishment, like moralizing high gods and political complexity, shows a high degree of phylogenetic scatter, meaning that cultures are often not similar to their relatives, suggesting either many independent gains or frequent losses or both. So, as for moralizing high gods, the ancestral state reconstruction gives between 25 and 50% probability of the ancestral state being broad supernatural punishment on almost every ancestral branch in the tree, despite only a third of sampled cultures showing this trait. An alternative interpretation is that the distribution of broad supernatural punishment across the tips of the phylogeny is shaped more by independent acquisitions and/or borrowings—but if this is true it casts doubt on the usefulness of the phylogenetic models underlying the analysis (Evans et al. 2021). Given that this class of phylogenetic methods can be prone to a high rate of false positives (Maddison and FitzJohn 2015), in an ideal world the conclusions would be subject to tests of sensitivity, power, and model adequacy (Hua and Bromham 2015).

The point here is not to critique the hypotheses linking these three traits, nor to suggest the phylogenetic analysis is flawed, but to highlight that the results of any phylogenetic study cannot be interpreted without reference to the assumptions made in the analysis (Bromham 2019). Whether this particular analysis is convincing evidence for links between political complexity, broad supernatural punishment, and moralizing high gods depends not on the association between their observed associations in living cultures (which is within the realms of chance co-occurrence) but on acceptance of the assumptions of the model concerning the way that both of these traits change along the phylogeny (in this case, as for many such analyses, it’s difficult to state those assumptions clearly, particularly because the method averages over five alternative models; Pagel and Meade 2006). Just as in the case studies of the phylogenetic histories of Australia discussed above, the believability of

the story depends not only on the results of the phylogenetic analysis but also on prior beliefs about the processes underlying the evolution of the traits. Sometimes the prior beliefs are obvious (such as calibration dates), and sometimes they are subtle (such as models of trait evolution underlying ancestral state reconstruction). These phylogenies are extremely useful for framing hypotheses and evaluating relative support for different ideas, but the answers from any phylogenetic study must be interpreted in light of the assumptions on which they rest.

Phylogenies Help Sort Meaningful Correlations from Incidental Association

Phylogenies play a critical role in testing hypotheses about evolution of human diversity, wherever conclusions are drawn by comparing observations from cultures that differ in the traits of interest. This is because we need to be able to separate meaningful association between cultural traits and the incidental relationships that are generated by inheritance from a common ancestor. To illustrate this point, we can consider two studies that show a correlation between aspects of environment and measures considered to reflect human intelligence: in the first instance, the frequency of Nobel prize winners in populations, and in the second, scores based on “IQ tests.”

In the first example, the number of Nobel prizes per country was found to be correlated with the per capita chocolate consumption rates (Messerli 2012). Given that cocoa has been implicated in improving cognitive function (Sokolov et al. 2013; Mastroiacovo et al. 2015; Barrera-Reyes et al. 2020), it is not unreasonable to suggest that chocolate consumption is somehow related to intellectual performance (or at least we hope it is). But the correlation across countries is not convincing evidence for this effect, because it is clear that the countries that have both high chocolate consumption and high Nobel success are geographically clustered and culturally related, being all northern European, predominantly Scandinavian, countries. In fact, any trait that is shared by these Scandinavian cultures will also correlate with both chocolate consumption and Nobel prizes, for example number of IKEA stores (Maurage et al. 2013). While we could hypothesize a causal link (such as “smart people like flat-pack furniture”), we are less inclined to interpret the link between Nobel prizes and IKEA stores as a meaningful relationship. While we can easily see here that counting each of these northern European cultures as an independent test of the association between intellectual performance, chocolate consumption, and IKEA stores inflates the perceived degree of association, the same problem is common to most cross-cultural studies (Roberts and Winters 2013; Bromham et al. 2018b).

The second example is the reported association between measures of IQ and parasite load (Eppig et al. 2010). Like the chocolate example, a reasonable hypothesis has been offered to explain the pattern: high parasite load might reduce investment in brain growth and function, thereby reducing cognitive performance. But, like the chocolate example, nearby cultures tend to be exposed to similar environments, and they tend also to be similar in cultural traits. This means related cultures tend to share many features, including average IQ scores (Gelade 2008) and parasite load (Guernier et al. 2004). This can lead to mistaken attribution of causality to indirect associations. Parasite load, like many biodiversity measures, has a latitudinal gradient and so will correlate with anything else with a latitudinal gradient, not only IQ (Kura 2013), but also population density, language diversity and gross domestic product (GDP) (Bromham et al. 2021b). IQ will also tend to correlate with anything else that correlates with latitude, not only parasite load but also mammal species richness (Bromham et al. 2018b), deforestation rates (Salahodjaev 2016), UV radiation (León and Burga León 2014), and GDP (Dickerson 2006; Lynn and Vanhanen 2012).

We can see an example of this correlation by proximity and relatedness by considering the “Darwinian gastronomy” hypothesis that spicy food has been favored in hot countries because it is protective against foodborne infection. Correlation of average number of spices per recipe with average temperature and parasite load has been taken as evidence that spicy food is a cultural adaptation to reducing risk of foodborne infection (Sherman and Billing 1999; Schaller and Murray 2010). But, because biodiversity, climate, and socioeconomic indicators show strong covariation, many other cultural variables have a significantly stronger correlation with spicy food than parasite load, including GDP and road traffic accidents (Bromham et al. 2021b). Which way do the causal arrows flow? Do parasites make you poor, prone to accidents and preferring spicy food? Does poverty scale with climate and therefore with parasites? Do parasites reduce your cognitive capacity, or does all biodiversity make you stupid (if the negative correlation between IQ and species richness is taken at face value: Bromham et al. 2018)? Correlations such as these can suggest hypotheses, but they cannot effectively test them. But we can use phylogenies to weight the relative explanatory potential of candidate causal hypotheses (Bromham et al. 2021b).

Phylogenies are a useful tool for protecting ourselves from over-interpreting associations that are due to the similarities generated by descent and proximity (Harvey and Purvis 1991). Related cultures are pseudo-replicates if they repeatedly sample the co-association of traits that were inherited together from a common ancestor, a statistical problem influencing all observations that are related

by descent (Bromham 2022). If you want to know whether having one trait (such as high parasite load) causes a change in another trait (such as average IQ), then including many different cultures or countries in a correlation of IQ against parasite load runs up against the problem of phylogenetic nonindependence, or Galton's problem: neighboring, related countries will tend to have similar values of cultural traits (such as IQ) and similar environmental features (like parasite load), creating the appearance of a relationship between the two, even if they are causally unconnected. But with a phylogeny, you are able to do the equivalent of an experiment: if the parasite load of a lineage increases, does its average IQ go down? By tracing the change in characteristics from shared ancestors to their current descendants, you can look for an association above and beyond the incidental association expected due to shared environment and heritage (Bromham 2017). Previously reported correlations between parasite load and sociosexuality (Schaller and Murray 2008), democracy (Murray et al. 2013), and language diversity (Fincher and Thornhill 2008) are weak or absent once phylogeny and spatial proximity are taken into account (Bromham et al. 2018b).

There are two broad ways that a phylogeny helps overcome the problem of phylogenetic non-independence. The first is that the phylogeny can be used to generate a matrix of covariance. This allows us to make an informed prediction of how similar cultures should be by historical inertia alone, and therefore allows us to detect relationships above and beyond that which comes “for free” from shared inheritance (e.g., PGLS, phylogenetic least squares regression) (Symonds and Blomberg 2014). The second approach is to use the phylogeny to choose comparisons between cultures that share a unique common ancestor (that is not shared with other such comparisons in the analysis), so that any differences between them in the traits of interest must have evolved independently in each lineage after they split from their shared ancestor (e.g., PIC, phylogenetically independent contrasts) (Garland Jr et al. 1992). The matrix of covariance can be extended to include the expected similarity due to distance (Hua et al. 2019), or even similarity due to contact and exchange between cultures (Bromham et al. 2021a). These approaches allow us to detect the signal of causal connections between traits above and beyond the correlations caused by the relationships between cultures.

Conclusion

So are phylogenies a good thing or a bad thing for understanding human history, evolution, and diversity? The answer is: they are a useful thing. But, like any source of evidence in understanding human evolution (or any aspect

of the evolutionary past), they are imperfect. Relying on phylogenetic evidence alone to understand cultural, linguistic, or cognitive evolution is risky, given the uncertainties. But phylogenies also allow us to peer into pasts that would otherwise be inaccessible to us, and give us a way of asking questions that is hard to match for power and usefulness. The special nature of phylogenies as historical inference is sometimes lost in the focus on statistical methods and large datasets.

Problems in reconstructing cultural evolution long-recognized in the social sciences have long-standing phylogenetic solutions in evolutionary biology. Yet the uptake of these useful methods has been inhibited by some misunderstandings of what phylogenetic methods can provide and what they require to be successful. Perhaps these two observations—that biologists sometimes have a limited understanding of the challenges of historical inference and that social scientists sometimes have an incomplete appreciation of the challenges of analyzing entities that are related by descent—suggest that evolutionary biologists and social scientists should spend more time talking to each other.

Acknowledgments With thanks to Ross Pain for his enthusiasm and energy in organizing the Cognitive Archaeology symposium at Australian National University in 2019; Lilly Yuen for supplying additional information on the hepatitis study; and to Quentin Atkinson, Claire Bowern, Simon Greenhill, Felicity Meakins, Ross Pain, Rhiannon Schembri, Kim Sterelny, Ray Tobler, and Keaghan Yaxley for helpful feedback.

Funding No funding was received for the preparation of this manuscript.

Open Access funding enabled and organized by CAUL and its Member Institutions

Availability of Data and Material All data and studies cited have been published.

Code Availability No code used.

Declarations

Conflict of Interest No relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright

holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Archibald JD (2014) Aristotle's ladder, Darwin's tree: the evolution of visual metaphors for biological order. Columbia University Press, New York
- Atkinson QD, Meade A, Venditti C, Greenhill SJ, Pagel M (2008) Languages evolve in punctuational bursts. *Science* 319(5863):588
- Bakker P, Borchsenius F, Levisen C, Sippola E (2017) Creole studies—phylogenetic approaches. John Benjamins Publishing, Amsterdam
- Balme J, O'Connor S, Fallon S (2018) New dates on dingo bones from Madura Cave provide oldest firm evidence for arrival of the species in Australia. *Sci Rep* 8(1):1–6
- Barrera-Reyes PK, de Lara JC-F, González-Soto M, Tejero ME (2020) Effects of cocoa-derived polyphenols on cognitive function in humans. Systematic review and analysis of methodological aspects. *Plant Foods Hum Nutr* 75(1):1–11
- Berlin B, Kay P (1969) Basic color terms: their universality and evolution. University of California Press, Berkeley
- Beyer R, Singarayer JS, Stock JT, Manica A (2019) Environmental conditions do not predict diversification rates in the Bantu languages. *Heliyon* 5(10):e02630
- Bouckaert RR, Bowern C, Atkinson QD (2018) The origin and expansion of Pama-Nyungan languages across Australia. *Nat Ecol Evol* 2(4):741–749
- Bowern C, Epps P, Gray R, Hill J, Hunley K, McConnell P, Zentz J (2011) Does lateral transmission obscure inheritance in hunter-gatherer languages? *PLoS ONE* 6(9):e25195
- Bowern C, Haynie HJ (2017) Reply to Nash: color terms are lost, despite missing data. *Proc Natl Acad Sci USA* 114(39):E8132–E8133
- Bowler JM, Johnston H, Olley JM, Prescott JR, Roberts RG, Shawcross W, Spooner NA (2003) New ages for human occupation and climatic change at Lake Mungo, Australia. *Nature* 421(6925):837–840
- Bromham L (2017) Curiously the same: swapping tools between linguistics and evolutionary biology. *Biology & Philosophy* 32(6):855–886
- Bromham L (2019) Six impossible things before breakfast: assumptions, models, and belief in molecular dating. *Trends Ecol Evol* 34(5):474–486
- Bromham L (2022) Solving Galton's problem: practical solutions for analysing language diversity and evolution. *Historical Linguistics* (in press)
- Bromham L, Dinnage R, Skirgård H, Ritchie AM, Cardillo M, Meakins F et al (2021a) Global predictors of language endangerment and the future of linguistic diversity. *Nat Ecol Evol*. <https://doi.org/10.1038/s41559-021-01604-y>
- Bromham L, Duchêne S, Hua X, Ritchie A, Duchêne D, Ho S (2018a) Bayesian molecular dating: Opening up the black box. *Biol Rev* 93(2):1165–1191
- Bromham L, Hua X, Cardillo M, Schneemann H, Greenhill SJ (2018b) Parasites and politics: why cross-cultural studies must control for relatedness, proximity and covariation. *Royal Soc Open Sci* 5(8):181100
- Bromham L, Hua X, Fitzpatrick TG, Greenhill SJ (2015) Rate of language evolution is affected by population size. *Proc Natl Acad Sci USA* 112: 2097–2102
- Bromham L, Skeels A, Schneemann H, Dinnage R, Hua X (2021b) There is little evidence that spicy food in hot countries is an adaptation to reducing infection risk. *Nat Hum Behav* 5:878–891
- Bryant D, Filimon F, Gray RD (2005) Untangling our past: languages, trees, splits and networks. In: Mace R, Holden CJ, Shennan S (eds) *The evolution of cultural diversity: a phylogenetic approach*. UCL Press, London, pp 67–83
- Buckley CD (2012) Investigating cultural evolution using phylogenetic analysis: the origins and descent of the Southeast Asian tradition of warp ikat weaving. *PLoS ONE* 7(12):e52064
- Clarkson C, Jacobs Z, Marwick B, Fullagar R, Wallis L, Smith M et al (2017) Human occupation of northern Australia by 65,000 years ago. *Nature* 547(7663):306–310
- Collard M, Shennan SJ, Tehrani JJ (2006) Branching, blending, and the evolution of cultural similarities and differences among human populations. *Evol Hum Behav* 27(3):169–184
- Crow JF (2001) The beanbag lives on. *Nature* 409(6822):771–771
- Currie A (2018) *Rock, bone, and ruin: an optimist's guide to the historical sciences*. MIT Press, Cambridge
- Currie TE (2013) Cultural evolution branches out: the phylogenetic approach in cross-cultural research. *Cross-Cultural Res* 47(2):102–130
- Dagan T, Martin W (2006) The tree of 1%. *Genome Biol* 7(10):1–7
- Darwin C (1859) *The origin of species by means of natural selection: or the preservation of favoured races in the struggle for life*. John Murray, London
- Da Silva SG, Tehrani JJ (2016) Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *Royal Soc open Sci* 3(1):150645
- Daval-Markussen A, Bakker P (2011) A phylogenetic networks approach to the classification of English-based Atlantic creoles. *Engl World-Wide* 32(2):115–146
- Dickerson RE (2006) Exponential correlation of IQ and the wealth of nations. *Intelligence* 34(3):291–295
- Doolittle WF (1999) Phylogenetic classification and the universal tree. *Science* 284(5423):2124–2128
- Dronamraju K (2011) *Haldane, Mayr, and beanbag genetics*. Oxford University Press, Oxford
- Duchêne DA, Hua X, Bromham L (2017) Phylogenetic estimates of diversification rate are affected by molecular rate variation. *J Evol Biol* 30(10):1884–1897
- Eppig C, Fincher CL, Thornhill R (2010) Parasite prevalence and the worldwide distribution of cognitive ability. *Proc Royal Soc B: Biol Sci* 277(1701):3801–3808
- Evans CL, Greenhill SJ, Watts J, List J-M, Botero CA, Gray RD, Kirby KR (2021) The uses and abuses of tree thinking in cultural evolution. *Philosophical Trans Royal Soc B* 376(1828):20200056
- Fillios MA, Taçon PS (2016) Who let the dogs in? A review of the recent genetic evidence for the introduction of the dingo to Australia and implications for the movement of people. *J Archaeol Science: Rep* 7:782–792
- Fincher CL, Thornhill R (2008) A parasite-driven wedge: infectious diseases may explain language and other biodiversity. *Oikos* 117(9):1289–1297
- Flowers JM, Hazzouri KM, Gros-Balthazard M, Mo Z, Koutroumpa K, Perrakis A et al (2019) Cross-species hybridization and the origin of North African date palms. *Proc Natl Acad Sci USA* 116(5):1651–1658
- François A (2014) Trees, waves and linkages: models of language diversification. In: Bowern C, Evans B (eds) *The Routledge handbook of historical linguistics*. Routledge, Abingdon, pp 161–189
- Friedman JE (2011) Anticipation in hereditary disease: the history of a biomedical concept. *Hum Genet* 130(6):705–714
- Garland T Jr, Harvey PH, Ives AR (1992) Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst Biol* 41(1):18–32
- Geisler H, List J-M (2013) Do languages grow on trees? The tree metaphor in the history of linguistics. In: Fangerau H, Geisler H, Halling T, Martin W (eds) *Classification and evolution in biology*,

- linguistics and the history of science. Concepts–methods–visualization. Franz Steiner Verlag, Stuttgart, pp 111–124
- Gelade GA (2008) The geography of IQ. *Intelligence* 36(6):495–501
- Gray RD, Atkinson QD, Greenhill SJ (2011) Language evolution and human history: what a difference a date makes. *Philosophical Trans Royal Soc B: Biol Sci* 366(1567):1090–1100
- Gray RD, Drummond AJ, Greenhill SJ (2009) Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323(5913):479–483
- Gray RD, Greenhill SJ, Ross RM (2007) The pleasures and perils of Darwinizing culture (with phylogenies). *Biol Theory* 2(4):360–375
- Gray RD, Watts J (2017) Cultural macroevolution matters. *Proc Natl Acad Sci USA* 114(30):7846–7852
- 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 SJ, Gray RD (2012) Basic vocabulary and Bayesian phylo-linguistics: issues of understanding and representation. *Diachronica* 29(4):523–537
- 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. *Proc Natl Acad Sci USA* 112(43):13296–13301
- Guernier V, Hochberg ME, Guégan J-F (2004) Ecology drives the worldwide distribution of human diseases. *PLoS Biol* 2(6):e141
- Haldane JBS (1964) A defense of beanbag genetics. *Perspect Biol Med* 7(3):343–360
- Hamilton MJ, Walker RS (2019) Nonlinear diversification rates of linguistic phylogenies over the Holocene. *PLoS ONE* 14(7):e0213126
- Harvey PH, Purvis A (1991) Comparative methods for explaining adaptations. *Nature* 351(6328):619–624
- Haynie HJ, Bowern C (2016) Phylogenetic approach to the evolution of color term systems. *Proc Natl Acad Sci USA* 113(48):13666–13671
- Heggarty P (2014) Prehistory through language and archaeology. In: Bowern C, Evans B (eds) *The Routledge handbook of historical linguistics*. Routledge, Abingdon, pp 598–626
- Hua X, Bromham L (2015) Phylometrics: an R package for detecting macroevolutionary patterns, estimating errors on phylogenetic metrics, and backward tree simulation. *Methods Ecol Evol* 7(7):806–810
- Hua X, Greenhill SJ, Cardillo M, Schneemann H, Bromham L (2019) The ecological drivers of variation in global language diversity. *Nat Commun* 10(1):2047
- Hull DL (1988) *Science as a process: an evolutionary account of the social and conceptual development of science*. University of Chicago Press, Chicago
- Jacques G, List J-M (2019) Save the trees: why we need tree models in linguistic reconstruction (and when we should apply them). *J Hist linguistics* 9(1):128–167
- Jordan FM (2011) A phylogenetic analysis of the evolution of Austronesian sibling terminologies. *Hum Biol* 83(2):297–321
- Jordan FM (2013) Comparative phylogenetic methods and the study of pattern and process in kinship. In: McConvell P, Keen I, Hendery R (eds) *Kinship systems: change and reconstruction*. University of Utah Press, Salt Lake City, pp 43–58
- Jordan P, O'Neill S (2010) Untangling cultural inheritance: language diversity and long-house architecture on the Pacific northwest coast. *Philosophical Trans Royal Soc B: Biol Sci* 365(1559):3875–3888
- Kalyan S, François A (2018) 5. Freeing the comparative method from the tree model: a framework for historical glottometry. *Senri Ethnological Studies* 98:59–89
- Kura K (2013) Japanese north–south gradient in IQ predicts differences in stature, skin color, income, and homicide rate. *Intelligence* 41(5):512–516
- Leach AD, Harris RB, Rannala B, Yang Z (2013) The influence of gene flow on species tree estimation: a simulation study. *Syst Biol* 63(1):17–30
- León FR, Burga León A (2014) Why complex cognitive ability increases with absolute latitude. *Intelligence* 46:291–299
- Levinson SC, Gray RD (2012) Tools from evolutionary biology shed new light on the diversification of languages. *Trends Cogn Sci* 16(3):167–173
- Lind J, Lindenfors P, Ghirlanda S, Lidén K, Enquist M (2013) Dating human cultural capacity using phylogenetic principles. *Sci Rep* 3(1):1–5
- Lynn R, Vanhanen T (2012) National IQs: a review of their educational, cognitive, economic, political, demographic, sociological, epidemiological, geographic and climatic correlates. *Intelligence* 40(2):226–234
- Mace R, Holden CJ (2005) A phylogenetic approach to cultural evolution. *Trends Ecol Evol* 20(3):116–121
- Mace R, Jordan FM (2011) Macro-evolutionary studies of cultural diversity: A review of empirical studies of cultural transmission and cultural adaptation. *Philosophical Trans Royal Soc B: Biol Sci* 366(1563):402–411
- Maddison WP, FitzJohn RG (2015) The unsolved challenge to phylogenetic correlation tests for categorical characters. *Syst Biol* 64(1):127–136
- Malaspinas A-S, Westaway MC, Muller C, Sousa VC, Lao O, Alves I et al (2016) A genomic history of Aboriginal Australia. *Nature* 538(7624):207–214
- Mastroiaco D, Kwik-Urbe C, Grassi D, Necozone S, Raffaele A, Pistacchio L et al (2015) Cocoa flavanol consumption improves cognitive function, blood pressure control, and metabolic profile in elderly subjects: the Cocoa, Cognition, and Aging (CoCoA) Study—a randomized controlled trial. *Am J Clin Nutr* 101(3):538–548
- Matthewson J (2011) Trade-offs in model-building: a more target-oriented approach. *Stud History Philos Sci Part A* 42(2):324–333
- Maurage P, Heeren A, Pesenti M (2013) Does chocolate consumption really boost Nobel award chances? The peril of over-interpreting correlations in health studies. *J Nutr* 143(6):931–933
- Mayr E (1963) *Animal species and evolution*. Harvard University Press, Cambridge
- McConvell P (2009) Loanwords in Gurindji, a Pama-Nyungan language of Australia. In: Haspelmath M, Tadmor U (eds) *Loanwords in the world's languages: a comparative handbook*. De Gruyter, Berlin, pp 790–822
- McConvell P, Bowern C (2011) The prehistory and internal relationships of Australian languages. *Lang Linguistics Compass* 5(1):19–32
- McInnis MG (1996) Anticipation: an old idea in new genes. *Am J Hum Genet* 59(5):973
- Meakins F, Pensalfini R, Zipf C, Hamilton-Hollaway A (2020) Lend me your verbs: verb borrowing between Jingulu and Mudburra. *Australian J Linguistics* 40(3):1–23
- Meakins F, Ennever T, Browne M, Browne M, Hamilton-Hollaway A (in press) Ngumpin-Yapa Languages. In: Bowern C (ed) *Handbook of Australian languages*
- Mesoudi A (2016) Cultural evolution: a review of theory, findings and controversies. *Evol Biol* 43(4):481–497
- Messerli FH (2012) Chocolate consumption, cognitive function, and Nobel laureates. *N Engl J Med* 367(16):1562–1564
- Moran NA, Jarvik T (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* 328(5978):624–627

- Morrison DA (2015) Review of: Aristotle's ladder, Darwin's tree: the evolution of visual metaphors for biological order. By J David Archibald *Systematic Biology* 64(5):892–895
- Murray DR, Schaller M, Suedfeld P (2013) Pathogens and politics: further evidence that parasite prevalence predicts authoritarianism. *PLoS ONE* 8(5):e62275
- Nash D (2017) Loss of color terms not demonstrated. *Proc Natl Acad Sci USA* 114(39):E8131–E8131
- Nelson-Sathi S, List J-M, Geisler H, Fangerau H, Gray RD, Martin W (2011) Networks uncover hidden lexical borrowing in Indo-European language evolution. *Proc Royal Soc B: Biol Sci* 278(1713):1794–1803
- O'Brien MJ, Boulanger MT, Buchanan B, Collard M, Lyman RL, Darwent J (2014) Innovation and cultural transmission in the American Paleolithic: phylogenetic analysis of eastern Paleoindian projectile-point classes. *J Anthropol Archaeol* 34:100–119
- Pagel M (2009) Human language as a culturally transmitted replicator. *Nat Rev Genet* 10(6):405–415
- Pagel M, Meade A (2006) Bayesian analysis of correlated evolution of discrete characters by Reversible-Jump Markov Chain Monte Carlo. *Am Nat* 167:808–825
- Paradis E (2014) An introduction to the phylogenetic comparative method. In: Garamszegi L (ed) *Modern phylogenetic comparative methods and their application in evolutionary biology*. Springer, New York, pp 3–18
- Paraskevis D, Magiorkinis G, Magiorkinis E, Ho SY, Belshaw R, Allain JP, Hatzakis A (2013) Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology* 57(3):908–916
- Pensalfini R, Meakins F (2019) Gender lender: noun borrowings between Jingulu and Mudburra in northern Australia. *J Lang Contact* 12(2):440–478
- Pian R, Archer M, Hand SJ (2013) A new, giant platypus, *Obdurodon tharalkooschild*, sp. nov. (Monotremata, Ornithorhynchidae), from the Riversleigh World Heritage Area, Australia. *J Vertebr Paleontol* 33(6):1255–1259
- Pietsch TW (2012) *Trees of life: a visual history of evolution*. Johns Hopkins University Press, Baltimore
- Rao V, Nanjundiah V (2011) J. B. S. Haldane, Ernst Mayr and the beanbag genetics dispute. *J Hist Biol* 44(2):233–281
- Ritchie AM, Ho SYW (2019) Influence of the tree prior and sampling scale on Bayesian phylogenetic estimates of the origin times of language families. *J Lang Evol* 4(2):108–123
- Roberts S, Winters J (2013) Linguistic diversity and traffic accidents: lessons from statistical studies of cultural traits. *PLoS ONE* 8(8):e70902
- Robinson LC (2015) The Alor-Pantar (Papuan) languages and Austronesian contact in East Nusantara. *Language Contact and Austronesian Historical Linguistics*. Pacific Linguistics, Canberra, p 19
- Romanes GJ (1897) *Darwin and after Darwin: an exposition of the Darwinian theory and a discussion of post-Darwinian questions*. Longmans, Green, London
- Ross MD (1988) *Proto Oceanic and the Austronesian languages of western Melanesia*. Pacific Linguistics, Australian National University, Canberra
- Rowland MJ (2018) 65,000 years of isolation in Aboriginal Australia or continuity and external contacts? An assessment of the evidence with an emphasis on the Queensland coast. *Anthropological Society of South Australia* 42:211–240
- Salahodjaev R (2016) Intelligence and deforestation: international data. *For Policy Econ* 63:20–27
- Schaller M, Murray DR (2008) Pathogens, personality, and culture: disease prevalence predicts worldwide variability in sociosexuality, extraversion, and openness to experience. *J Personal Soc Psychol* 95(1):212
- Schaller M, Murray DR (2010) Infectious diseases and the evolution of cross-cultural differences. *Evolution, culture, and the human mind*:243–256
- Schleicher A (1863) *Die darwinsche Theorie und die Sprachwissenschaft*. Hermann Böhlau, Weimar
- Shafir A, Azouri D, Goldberg EE, Mayrose I (2020) Heterogeneity in the rate of molecular sequence evolution substantially impacts the accuracy of detecting shifts in diversification rates. *Evolution* 74(8):1620–1639
- Sherman PW, Billing J (1999) Darwinian gastronomy: why we use spices: spices taste good because they are good for us. *Bioscience* 49(6):453–463
- Slack MJ, Law WB, Gliganic LA (2020) The early occupation of the Eastern Pilbara revisited: new radiometric chronologies and archaeological results from Newman Rockshelter and Newman Orebody XXIX. *Q Sci Rev* 236:106240
- Sokolov AN, Pavlova MA, Klosterhalfen S, Enck P (2013) Chocolate and the brain: neurobiological impact of cocoa flavanols on cognition and behavior. *Neurosci Biobehavioral Reviews* 37(10):2445–2453
- Symonds MR, Blomberg SP (2014) A primer on phylogenetic generalised least squares. In: Garamszegi LZ (ed) *Modern phylogenetic comparative methods and their application in evolutionary biology*. Springer, New York, pp 105–130
- Tadmor U (2009) Loanwords in the world's languages: findings and results. In: Haspelmath M, Tadmor U (eds) *Loanwords in the world's languages. A comparative handbook*. De Gruyter, Berlin, pp 55–75
- Tehrani J, Collard M (2002) Investigating cultural evolution through biological phylogenetic analyses of Turkmen textiles. *J Anthropol Archaeol* 21(4):443–463
- Temkin I, Eldredge N (2007) Phylogenetics and material cultural evolution. *Curr Anthropol* 48(1):146–154
- Tiley GP, Poelstra JW, dos Reis M, Yang Z, Yoder AD (2020) Molecular clocks without rocks: new solutions for old problems. *Trends Genet* 36(11):845–856
- Tobler R, Hermisson J, Schlötterer C (2015) Parallel trait adaptation across opposing thermal environments in experimental *Drosophila melanogaster* populations. *Evolution* 69(7):1745–1759
- Tobler R, Rohrlach A, Soubrier J, Bover P, Llamas B, Tuke J et al (2017) Aboriginal mitogenomes reveal 50,000 years of regionalism in Australia. *Nature* 544(7649):180
- Watts J, Greenhill SJ, Atkinson QD, Currie TE, Bulbulia J, Gray RD (2015) Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proc Royal Soc B: Biol Sci* 282(1804):20142556
- Welch JJ, Bromham L (2005) Molecular dating when rates vary. *Trends Ecol Evol* 20:320–327
- Windram HF, Charlston T, Howe CJ (2014) A phylogenetic analysis of Orlando Gibbons's *Prelude in G*. *Early Music* 42(4):515–528
- Yuen LKW, Littlejohn M, Duchêne S, Edwards R, Bukulatjpi S, Binks P et al (2019) Tracing ancient human migrations into Sahul using hepatitis B virus genomes. *Mol Biol Evol* 36(5):942–954
- Zhang H, Ji T, Pagel M, Mace R (2020) Dated phylogeny suggests early Neolithic origin of Sino-Tibetan languages. *Sci Rep* 10(1):1–8