



**HAL**  
open science

## The Family Tree model

Thomas Pellard, Robin Ryder, Guillaume Jacques

► **To cite this version:**

Thomas Pellard, Robin Ryder, Guillaume Jacques. The Family Tree model. Ledgeway, Adam; Aldridge, Edith; Breitbarth, Anne; Kiss, Katalin É.; Salmons, Joseph; Simonenko, Alexandra. The Wiley Blackwell companion to diachronic linguistics, Wiley Blackwell, 2026. <hal-04571512v2>

**HAL Id: hal-04571512**

**<https://hal.science/hal-04571512v2>**

Submitted on 21 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.




Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# The Family Tree model

Thomas Pellard\*, Robin J. Ryder\*\*, and Guillaume Jacques\*\*\*

\*  0000-0001-6189-1677 CRLAO, EHESS, CNRS, Inalco [thomas.pellard@cnrs.fr](mailto:thomas.pellard@cnrs.fr)

\*\*  0000-0002-7467-0467 Department of Mathematics, Imperial College London

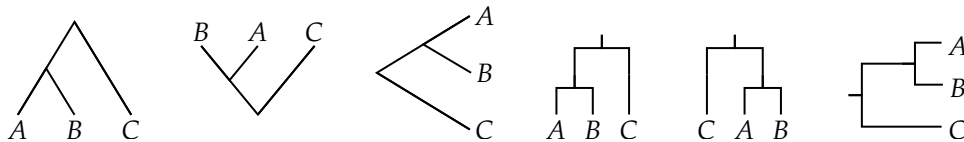
\*\*\*  0000-0003-1220-2323 CRLAO, EHESS, CNRS, Inalco & EPHE, PSL University

Preprint 2025-01-21, to appear in Adam Ledgeway, Edith Aldridge, Anne Breitbarth, Katalin É. Kiss, Joseph Salmons & Alexandra Simonenko (eds.). 2025. *The Wiley Blackwell companion to diachronic linguistics*. 5 vols. Wiley Blackwell.

## 1 Defining the Tree model

The Family Tree model represents the evolutionary history and relationships of languages as a tree resembling a genealogical tree (Figure 1). Though metaphors for historical relationships between languages in terms of common origin, kinship, and genealogy go back well beyond the modern times (Robins 1973, List et al. 2016), our current Tree model descends directly from the *Stammbaum* of Schleicher (1861–1862), who helped make historical and comparative linguistics a science (Fox 1995: 23–27), and it has remained at the heart of the discipline ever since. In the last decades, the models and methods of phylogenetics, the study of evolutionary history and relationships between groups of biological organisms, have been adapted to the study of languages and have helped enhance the traditional Tree model, leading to a new thrust in the study of linguistic evolution and to the emergence of the field of *phylolinguistics* (Gray, Greenhill & Atkinson 2013, Greenhill, Heggarty & Gray 2020). We will focus here on those recent developments rather than on the traditional model.

Within the Tree model, processes of historical changes and relationships are inferred from patterns in the distribution of observed similarities and differences. It is thus necessary to distinguish similarities that are due to shared ancestry (*homologies*), such as cognate words, from those due not to common descent but to independent evolution (*homoplasies*)



**Figure 1:** Equivalent representations of the same tree  $((A, B), C)$ , where  $A$  and  $B$  are more closely related than to  $C$  and thus form a subgroup

or contact (§2). However, homologies are not all equally informative for reconstructing linguistic history. Only shared innovations (*synapomorphies*; §3.1) are indicative of a shared history and constitute evidence that languages form a subgroup that is more closely related and has diverged more recently than their other relatives. The retention of features inherited without change from the common ancestor of all languages in a family (*symplesiomorphies*) cannot, by definition, constitute evidence of the historical diversification of that family, and such shared archaisms cannot define subgroups among related languages. This principle was first proposed in linguistics by Neogrammarians such as Leskien (1876), long before it was independently rediscovered in biology by Hennig (1950, 1979).

The Tree model thus explains the distribution of homologous characters within a language family by formulating hypotheses of diachronic changes in these characters, and languages which share such changes are inferred to be more closely related than languages that do not. For example, if two languages *A* and *B* share some innovative features that are not shared with a third language *C*, it is inferred that this is likely to be because those features arose after the split of *C* from the common ancestor of the three languages, but before the split of *A* and *B*, i.e. *A* and *B* are more closely related and share a more recent common ancestor than with *C*. This can then be summarized in the form of a parenthetical notation  $((A, B), C)$  or, more practically, as a tree diagram representing which languages are more closely related and the order of their divergence (Figure 1). This tree is interpreted as a *phylogeny*, i.e. the evolutionary history of the languages considered.

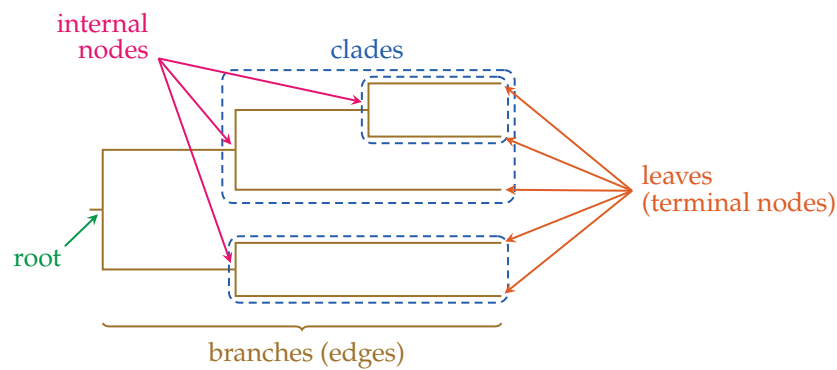
When distributions of different characters are incongruent, i.e. they disagree on which languages are to be grouped together and thus yield discordant trees, e.g.  $((A, B), C)$  vs.  $(A(B, C))$ , special processes are inferred to have occurred (§2). Besides independent parallel changes (§2.3), reversals, and incomplete lineage sorting (§2.4), *reticulation* processes where features are not transmitted from ancestor to descendant but transferred between divergent languages are noteworthy since they contradict the assumptions of the Tree model and cannot be represented on a tree but require a network (§2). However, reticulation processes can only be defined by opposition to tree-like descent processes, and they can only be identified and studied by assuming the existence of a tree.

Despite repeated criticisms and calls for alternative models, the tree remains the central metaphor and main model of the evolutionary history of languages (§1.4). Linguists have been fully aware of the limitations and problems of the Tree model (§2) ever since Schleicher, but it has proven useful not just for classifying languages, but also for understanding language diversification and acquiring new insights into evolutionary patterns and processes in language (§4).

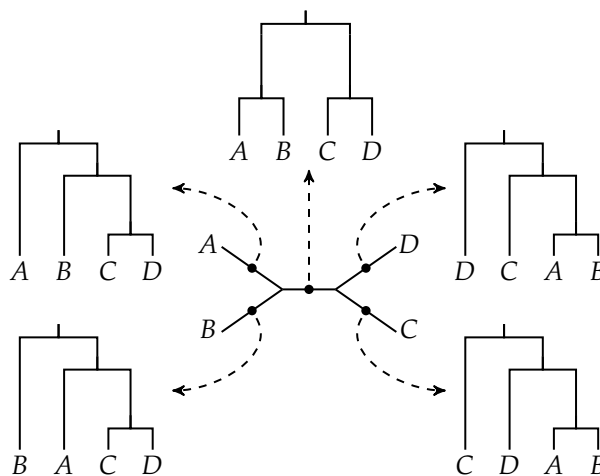
### 1.1 What is a tree?

The conceptual Tree model sketched above can be further refined by mathematical formalization, enabling precise and rigorous descriptions and inference procedures to be formulated. A tree, or *dendogram*, can be formally defined as a connected graph with no cycles. Each

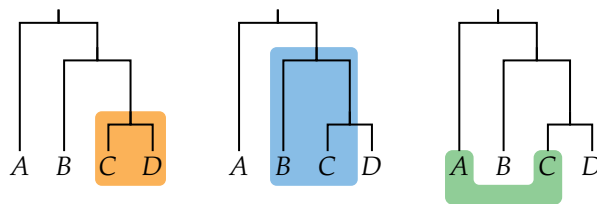
*edge*, or *branch*, represents a *lineage*, a sequence of languages related by ancestry-descent relationships (Fig. 2). Typically, observed languages, modern or ancient, are the *leaves* (also *tips* or *terminal nodes*) of the tree and have no children, while *internal nodes* represent events of divergence and are associated with the hypothetical most recent common ancestors of different lineages, i.e. unobserved proto-languages. A tree also usually has a *root* node that has no parent and represents the common ancestor of all languages in the tree. In contrast, an unrooted tree lacks a root node and does not directly represent ancestor-descendant relationships. It only provides information about *possible* subgroups, which can be found by alternatively placing a root node on every edge (Figure 3).



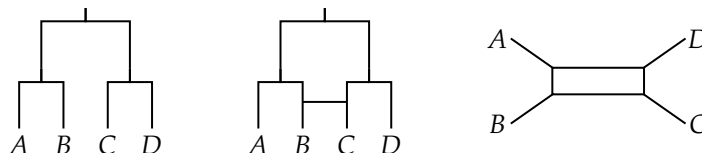
**Figure 2:** A tree and its different parts



**Figure 3:** The five different possible bifurcating rooted trees that can be obtained from a four-language unrooted tree by placing the root node on different edges; this unrooted tree excludes 10 other bifurcating topologies such as  $((A, C), (B, D))$ ,  $((A, D), (B, C))$ ,  $(A, (C, (B, D)))$ , etc., as possible rooted trees



**Figure 4:** Examples of a monophyletic (left), paraphyletic (center), and polyphyletic (right) group



**Figure 5:** A rooted tree (left), a rooted evolutionary network (center), and an unrooted data-display network (right)

Cutting a single branch on a tree isolates a *clade*, a subgroup that includes an ancestor and all its descendants (Figure 2). A clade is a *monophyletic* group that differs from both *paraphyletic* groups, which include an ancestor and some but not all its descendants, and *polyphyletic* groups, which include some descendants of a common ancestor but not the ancestor itself (Figure 4). Only monophyletic groups are natural groups that are meaningful from an evolutionary point of view, as they are groups in which languages are more closely related to each other than to any language outside the group. (Hennig 1979: 73).

The branching structure of the tree is called the *topology*. A tree is often constrained for practical reasons to be *binary*, also *bifurcating* or *dichotomous*: each internal node has exactly two child nodes, though a trifurcation can be represented as two bifurcations separated by a very short branch. A tree that only represents a topology is a *cladogram*, while a *phylogram* has branches that differ in length according to the amount of change attributed to the branch. When branch lengths are proportional to time, this makes the tree a *chronogram*, which can be absolute or relative. In a dated tree, leaf languages have age 0, unless they represent ancient languages.

A tree differs from a network in that in a tree each node has only one parent, whereas in a network a node may have multiple parents (Figure 5), which allows non-tree-like, reticulate evolutionary events to be represented (§2). It is important to note that many commonly used network types, such as neighbor-nets (Bryant & Moulton 2004), simply summarize graphically both tree-like and non-tree-like patterns in the data, with no prior assumptions about the causes of conflict in the data. Such data-display networks differ from both trees and evolutionary networks in that their internal nodes and edges may not represent ancestors and historical relationships. There is no guarantee that they represent evolutionary history, and they should therefore not be considered evolutionary models (Morrison 2011, Jacques & List 2019).

## 1.2 *What is a model?*

A model is a simplified representation of a complex phenomenon, usually explicitly and precisely formalized as a mathematical object, which captures its fundamental properties while remaining easy to manipulate, analyze, and understand. Models such as the Tree model are sometimes criticized for their lack of realism, accuracy and comprehensiveness as representations of linguistic evolution (§1.4), but this is the very nature of all scientific models in general, and this lack of faithfulness to reality can actually be useful and desirable (Velasco 2012, Elgin 2017, Potochnik 2017), like a schematic map is often more useful than a realistic depiction of the actual geography. A hackneyed aphorism in statistics, attributed to George Box (1979), is that “all models are wrong, but some are useful”; we might say instead that all models are wrong, but that this is not a bug: it is a feature.

A model usually attempts to generalize and represent categories rather than actual cases, it is thus bound to be partial and more abstract than the concrete individual phenomena it aims at representing. It is also usual for a model to only include core factors, leaving out those properties that may be safely considered negligible, in the sense that they can be safely distorted or ignored without significantly affecting our predictions or understanding. This facilitates both mathematical computations and the attribution of causes to observed outcomes. Actually, the most complex model is not necessarily the best one, quite the contrary: a model with too many parameters is not only unparsimonious, it also easily becomes overwhelmingly complex and intractable, and it thus fails to help our understanding of the relevant phenomenon. This is why for example the Tree model deliberately abstracts away from the details of microevolution at the level of individual speakers and communities. Much of the work in modelling is to build a model that is complex enough to capture the key aspects of the data, yet simple enough to be useful in practice.

A model also often describes ideal types that do not exist as such in the world, like ideal gases, perfectly rational agents, planet-like atoms, infinite populations, etc. Although this can seem counterintuitive, without regard for whether it is true or false that languages that only diverge without any further contact actually exist, a model that assumes such ideal cases can actually help infer that contact events have occurred. The Tree model assumes that languages only evolve by bifurcating divergence, and it thus predicts that all characters in a data set will have compatible distributions and can be represented by a single tree without any conflict. By making this unrealistic assumption, when the prediction does not hold, we can infer that evolution may not have proceeded by strict divergence (§2). In this sense, if an imperfect and inaccurate model is a fiction, it nevertheless may be more useful and virtuous than a more realistic and complex but less constrained model that fails to make precise predictions, except that anything goes.

## 1.3 *What is the Tree model a model of?*

The Tree model aims to represent the historical diversification of languages, assuming that linguistic features are inherited with modifications over time and that related languages

descend from a single common ancestor by divergence. It deliberately does not represent all the complexity of linguistic change in full detail, nor does it aim to represent all cases. It should thus not be misunderstood to represent phenomena that it was never meant to depict.

In particular, the Tree model is not intended to study cases where no such single ancestor can be identified, as in the case of mixed languages (§2.2), nor linguistic microevolution, i.e. relationships between individuals, which are usually irreducibly reticulate. It is best suited to cases of well-differentiated and geographically isolated languages, but usually ill-suited to cases of mutually intelligible dialects. However, in such cases, the fit of the model to the data can still be quantified, and phylogenetic methods can thus provide interesting insights (§3.6).

#### 1.4 *Competing models of linguistic diversification: Trees and waves*

The Tree model is the main model used to study language evolution, and linguists have always recognized its limitations that are due to the simplifications, abstractions and idealizations it makes (§1.2). These limitations have sometimes been regarded as unacceptable deficiencies, and alternative models were called for (Schmidt 1872, Schuchardt 1900, Bloomfield 1933, Trubetzkoy 1939, Dixon & Aikhenvald 2001, François 2015, Kalyan & François 2018). One of the few alternatives proposed is the Wave model (*Wellentheorie*), which views linguistic diversification as the result of series of independent waves of linguistic change that propagate across space with different ranges and speeds, affecting languages and dialects in contact (Schmidt 1872, Schuchardt 1900, François 2015, Kalyan & François 2018; WBCDL060 *Wave model: Diffusion, convergence, and differentiation*). One motivation for this model stems from the observation that isoglosses on linguistic maps tend to intersect, while under the Tree model they are expected to form bundles that do not intersect.

The Wave model is claimed to be more realistic since languages rarely if ever diverge by neat splitting without subsequent contact as assumed in the Tree model, but the Wave model also makes simplifying unrealistic assumptions. For example, it ignores social factors that are important in linguistic change and its diffusion, such as status and power, or the details of geography such as topography (Winter 1973, Britain 2013). Moreover, wave diagrams, like isogloss maps, do not model the process of language change as the Tree model does but only display a static summary of the results of diffusion, and they thus lack a temporal dimension (Anttila 1989: 306, Geisler & List 2013, Kalyan & François 2019).

Although often presented as a complement or even substitute for the Tree model, the Wave model has rarely been used in practice to model language evolution, probably because it has usually remained a mere conceptual metaphor (Hartmann 2023: 7, Geisler & List 2022). Without a precisely formulated model, all linguistic characters (sound changes, lexical and morphological innovations) are treated on a par without considering that different types of characters can evolve differently and require distinct evolutionary models (Greenhill et al. 2010). No attempt is made either to identify the various sources of incongruence

in the distribution of shared innovations, not all of which are due to diffusion (§2). Like the traditional method of tree inference, but not the modern phylogenetic methods (§3.2; [Brower & Schuh 2021](#): 113–141), the Wave model also necessitates to determine beforehand which features constitute innovations, which is usually not straightforward (§3.1), except in the case of particular spatial configurations.

An improvement on the traditional Wave model is Historical Glottometry ([François 2015](#), [Kalyan & François 2018](#)), which proposes several quantitative measures of subgroup support, but not a formal model of evolutionary processes. It actually does not challenge the Tree model once incomplete lineage sorting (§2.4) has been taken into account ([Kalyan & François 2019](#)).

Attempts to provide a formal, mathematical, version of the Wave model are rare, but both [Sankoff & Sankoff \(1980\)](#) and [Hartmann \(2023\)](#) agree that the Wave model can complement but not replace the Tree model. This concurs with the broad consensus among linguists that trees are necessary and that wave-like diffusion exists but is not the main mechanism of language evolution ([Gleason 1961](#), [Winter 1973](#), [Ross 1997](#), [Gąsiorowski 1999](#), [Haspelmath 2004](#), [Labov 2007](#), [Campbell & Poser 2008](#): 326–329, [Gray, Bryant & Greenhill 2010](#), [Joseph 2012](#), [Jacques & List 2019](#), [Evans et al. 2021](#), [Carling, Cathcart & Round 2021](#), [Geisler & List 2022](#)).

As in biology ([Franklin-Hall 2010](#)), the relevant questions are not whether tree models should be used but when they can, and how good a model is a tree for a given dataset ([Bowern 2018](#), [Auderset et al. 2023](#), [Hartmann 2023](#): 217–218). Despite the recurring misconception that linguists working with the Tree model neglect the importance of contact and areal diffusion, phylogenetic methods based on the Tree model have proven apt to identify and quantify both tree-like and non-tree-like aspects of language family’s evolution, and thus to study contact events and dialect continua (§3.6; [List et al. 2014](#), [Kelly & Nicholls 2017](#), [Kaiping & Klamer 2022](#), [Neureiter et al. 2022](#), [Auderset et al. 2023](#), [Hartmann 2023](#)).

## 2 Disentangling reticulation and the sources of discordance

Criticisms towards the Tree model often focus on the existence of conflicting data, i.e. characters whose distribution is incongruent with that of other characters ([Schmidt 1872](#), [Schuchardt 1900](#), [Bloomfield 1933](#), [Bonfante 1946](#), [François 2015](#), [Kalyan & François 2018](#)). Such data imply different tree topologies and thus cannot be adequately represented on a single tree. Phylogenetic data almost always contain such conflicting data, but this does not imply that the Tree model is inadequate nor that phylogenetic methods simply ignore such conflicts. As in biology ([Baum & Smith 2013](#): 161), “[m]ost of the interesting issues in phylogeny reconstruction are in how to resolve these conflicts” ([Felsenstein 1988](#): 254).

Not all conflicting data originate from reticulate, non-tree-like, evolution in which features are transferred between distinct lineages ([Morrison 2011](#): 44). Among the many sources of conflicting data, only true reticulation processes such as horizontal transfer (§2.1) and hybridization (§2.2) contradict the assumptions of the Tree model, and they should be strictly

distinguished from other processes such as parallel evolution (§2.3) and incomplete lineage sorting (§2.4). Ordinary, data-display networks are unable to make this distinction between actual reticulated evolution and “pseudo-reticulation” processes that produce data that only have the appearance of reticulation. An evolutionary network that only represents true reticulation events requires a phylogenetic tree as its backbone (Morrison 2011: 47, Lehtinen et al. 2014, List et al. 2014). It is precisely by logically excluding the possibility of conflicting data that the Tree model is able to highlight them as such for further study (Morrison 2011: 44), and recent phylogenetic models attempt to identify and represent reticulation (Kelly & Nicholls 2017, Neureiter et al. 2022).

### 2.1 Horizontal transfer: Borrowings

Languages more often than not influence each other and exchange features through contact, either structural features (phonological or morphosyntactic structures) and processes (sound changes), or linguistic substance (words, morphs). While the common practice is to base linguistic phylogenies on cognate lists and not on similarities in structural features (Campbell & Poser 2008, Bowerman 2018, Greenhill, Heggarty & Gray 2020), which can easily be due to contact or parallel changes (§2.3), loanwords are a typical source of conflict in phylogenies. Nevertheless, several studies have shown that phylogenetic methods are able to infer trees despite the existence of horizontal transfer, and that such reticulation does not invalidate the Tree model (Greenhill, Currie & Gray 2009, Bowerman et al. 2011, Barbañon et al. 2013). Reasonable amounts of horizontal transfer, even undetected, will only make the phylogenetic signal less clear but not undetectable, and new models are being developed for directly identifying borrowings within the data (Kelly & Nicholls 2017, Neureiter et al. 2022).

In the case of related languages, distinguishing between true, inherited, cognates from borrowings (from one language to the other or from a third language into both) can be difficult. Loanwords may however exhibit specific sound correspondences that set them apart from cognates and allow for their identification. For instance, Proto-Indo-European *\*p-* corresponds to either *h-* or zero in word-initial position in Armenian (first three etyma in Table 1). By contrast, in borrowings from Iranian, initial *p-* (unchanged from Proto-Indo-European *\*p-*) corresponds to *p-* in Armenian, since this layer was borrowed *after* the sound change *\*p- > h-*, and was therefore unaffected by it. These correspondences provide an easily applied criterion to distinguish between cognates and loanwords for words whose onset consonant originates from Proto-Indo-European *\*p-*.

Differences in the sound correspondences can serve to identify loanwords, but it is not always possible to find such differences. When a form has been relatively unaffected by sound changes, it can be impossible to determine whether it was inherited from the proto-language or whether it was borrowed at a later stage from a closely related variety. This is the case of the Armenian forms in Table 2, whose sound correspondences are the same in the inherited layer and in the loanwords from Iranian (Hübschmann 1897: 16–17).

**Table 1:** Correspondences of Proto-Indo-European initial *\*p-* in Armenian in native words and borrowings from Iranian (the main Iranian donor to Armenian, Parthian, is poorly attested, and other Iranian languages are given here for comparison instead)

Armenian	Meaning	PIE	Iranian (Avestan)
<i>hayr</i>	‘father’	<i>*ph<sub>2</sub>ter-</i>	<i>pitar-</i>
<i>hing</i>	‘five’	<i>*penk<sup>w</sup>e</i>	<i>pañca</i>
<i>otn</i>	‘foot’	<i>*pod-m</i>	<i>pad-</i>
<i>pah</i>	‘guard’		<i>pāθra-</i>
<i>partēz</i>	‘garden’		<i>pairidaeza-</i>
<i>pēs</i>	‘manner’		<i>paesa-</i>

**Table 2:** Armenian words which cannot be conclusively demonstrated to be either borrowings from Iranian or inherited words on the basis of their sound correspondences

Armenian	Meaning	Proto-Iranian
<i>naw</i>	‘boat’	<i>*nāu-</i>
<i>mēg</i>	‘mist’	<i>*maiga-</i>
<i>mēz</i>	‘urine’	<i>*maiza-</i>
<i>sar</i>	‘head’	<i>*sarah-</i>
<i>ayrem</i>	‘burn’	<i>*Haid-</i>

Such undetectable loanwords are not the only problem with using sound correspondences as a criterion for distinguishing between loanwords and cognates. When two related languages share an important number of cognates, bilingual speakers can develop an intuition of the phonological correspondences between the two languages, and apply them to newly borrowed words, a phenomenon known as *loan nativization* (Aikio 2007). For instance, Northern Saami *báhppa* ‘priest’ presents the same vowel correspondence *a :: i* with Finnish *pappi* ‘priest’ as etyma inherited from Proto-Finno-Ugric such as ‘name’ Northern Saami *namma* :: Finnish *nimi* (< *\*nimi*). However, Finnish *pappi* ‘priest’ is borrowed from Russian *pop* ‘priest’, and Saami *báhppa* ‘priest’ was not therefore inherited from Finno-Ugric but borrowed from Finnish. It has entered the Saami language too recently to have undergone the series of regular sound changes leading from Proto-Finno-Ugric *\*i* to Saami *a*, and its vocalism must have been modelled on that of native words such as ‘name’. This sound correspondence cannot therefore be used to distinguish between native and borrowed words.

Although both undetectable and nativized loanwords are problematic, they remain extreme cases. Usually, discriminating between cognates and borrowings is possible once the historical phonology of the languages investigated has been elucidated, leaving only a small residue of undecidable cases.

## 2.2 Hybridization and mixed languages

A real challenge to the Tree model is the existence of mixed languages with split ancestry (Bakker 2017, WBCDL043 Mixed languages), such as Michif (verbs and verbal morphology from Cree, nouns and nominal morphology from French) or Media Lengua (grammar from Quechua, lexicon from Spanish). Mixed languages are exceptional, and given their rarity, the Tree model remains generally applicable as long as such cases are correctly identified.

Mixed languages could be included in phylogenetic analyses by treating their different components separately. For example, the French component of Michif could be used in a phylogeny of Indo-European, and its Cree component in a phylogeny of Algonquian. However, an analysis based on the Tree model should not include both Indo-European and Algonquian languages, with Michif serving as a bridge between the two.

## 2.3 Parallel innovations

Although linguistic evolution is largely unpredictable, it does not follow entirely random processes, and historical and comparative linguistics has long identified recurrent patterns of change (see, for example, Hagege & Haudricourt 1978, Traugott & Dasher 2001, Kümmel 2007, Kuteva et al. 2019; WBCDL006 Possible and impossible changes; WBCDL016 Diachronic typological universals; WBCDL074 Phonological universals; WBCDL089 Morphological universals and tendencies; WBCDL114 Syntactic universals; WBCDL136 Semantic universals). It is therefore not surprising that the same changes occur independently in different languages, and this kind of parallel change is a major source of incongruence between characters, though it differs from reticulation proper. Parallel innovations are most conspicuous in the case of structural features (Campbell & Poser 2008, Greenhill et al. 2010, 2017), including phonological and syntactic features, and this is the main reason why phylogenies are usually inferred from lexical data (Bowern 2018, Greenhill, Heggarty & Gray 2020).

However, lexical innovations are also not free from parallel evolution (Chang et al. 2015). For example, several Indo-European languages share an innovative etymon *\*lowksneh<sub>2</sub>* for ‘moon’ derived from the root *\*lewk-* ‘bright, white’: Latin *lūna*, Old Slavonic *luna*, and Armenian *lusin* (Buck 1949: 54–55). However, this most likely does not constitute an innovation shared by Italic, Slavic and Armenian, but rather a parallel semantic evolution from a noun meaning ‘bright one’ (‘moonshine’) to ‘moon’. Similar examples of this semantic evolution are attested in unrelated language families, notably the colexification of ‘bright’ and ‘moon’ in Austronesian languages (Proto-Austronesian *\*bulaN*; Blust & Trussel 2010).

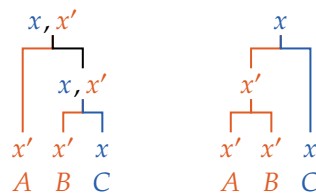
## 2.4 Incomplete lineage sorting

Another important, but seldom noted, process that obscures language phylogenies arises from the transmission of *polymorphisms*, i.e. variation. It is expected that ancestral languages, like any natural language, are subject to variation due to e.g. register or dialectal differences. If the same variant is lost independently in different descendants belonging to different

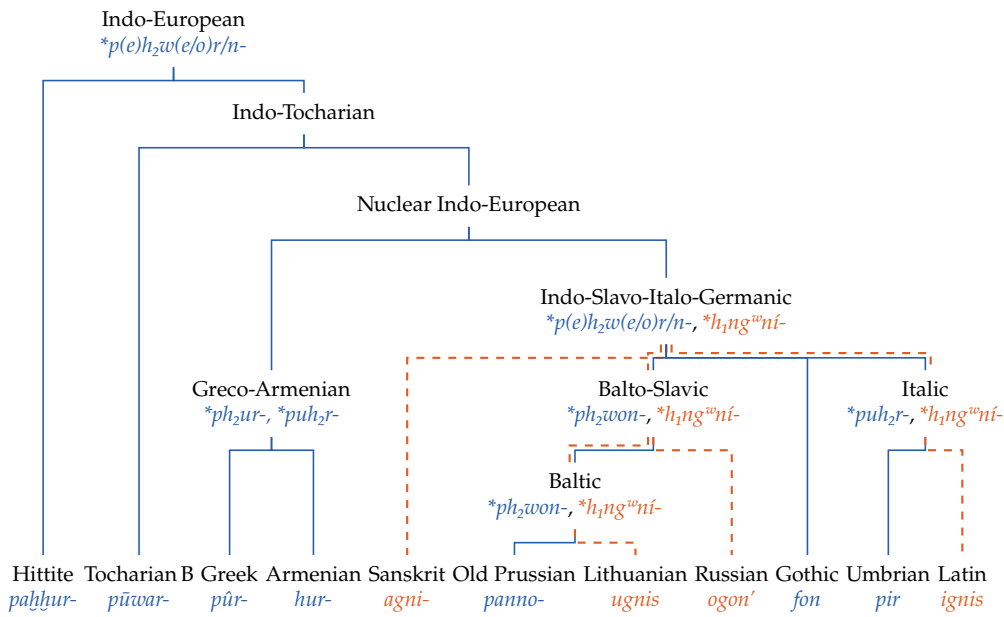
subgroups, this creates an incongruence with other characters and obscures the phylogeny. This phenomenon, known in evolutionary biology as *incomplete lineage sorting* (Maddison 1997, Galtier & Daubin 2008), can explain some cases of conflicting isoglosses (Jacques & List 2019) and explains how *etymon trees* differ from *language trees*, just as *gene trees* differ from *species trees* (Figure 9; Maddison 1997). In the case of incomplete lineage sorting, an innovation has already occurred in the proto-language but is not yet fixed, i.e. it has not yet replaced the other original variant in the whole speech community.

Figure 6 illustrates this phenomenon with a fictitious example. Suppose that the common ancestor of three languages whose actual phylogeny is represented by the left-hand tree in Figure 6 has two different variants for a given character,  $x$  and a more innovative  $x'$ . Each language has preserved a single variant, either  $x$  or  $x'$ , but since  $x'$  is the innovative variant, the distribution of the variants  $x, x'$  incorrectly suggests that languages  $A$  and  $B$  form a clade with their ancestor in which the change  $x > x'$  occurred, and that  $C$  does not belong to this clade and has not undergone any change, as shown in the right-hand tree of Figure 6. However, the proto-language already possessed both variants that were preserved in the common ancestor of  $B$  and  $C$ , and  $x$  was independently lost in both  $A$  and  $B$ , while  $x'$  was lost in  $C$ . In this example, the distribution of  $x$  and  $x'$  obscures the actual phylogeny.

The case of the two Indo-European etyma for ‘fire’,  $*p(e)h_2w(e/o)r/n-$  and  $*h_1ng^wni-$  (Wodtko, Irslinger & Schneider 2008: 340–341), provides an actual example of incomplete lineage sorting. If we place the reflexes of these two etyma on the tree of Chang et al. (2015) including only clades with more than 90% posterior probability, we observe that only  $*p(e)h_2w(e/o)r/n-$  occurs in the group comprising Anatolian, Tocharian and Greco-Armenian, and it is thus a good candidate for the original word for ‘fire’ in the proto-language. Within the Indo-Slavo-Italo-Germanic group however, the distribution of  $*h_1ng^wni-$  is not straightforwardly accounted for by the phylogeny, since Baltic and Italic have reflexes of both  $*h_1ng^wni-$  and  $*p(e)h_2w(e/o)r/n-$ . Using the presence of  $*h_1ng^wni-$  to define a clade comprising Sanskrit, Lithuanian, Russian and Latin, but excluding Old Prussian and Umbrian would conflict with the bulk of our knowledge on the history of these languages. Instead, we must assume that both etyma existed and were still in competition in Proto-Baltic (and hence Proto-Balto-Slavic) and in Proto-Italic, and that one of the two was independently lost in the daughter languages. The etymon  $*h_1ng^wni-$  has been innovated after the split of Greco-Armenian, but  $*p(e)h_2w(e/o)r/n-$  did not disappear, and the two may have been maintained as quasi-synonyms.



**Figure 6:** Incomplete lineage sorting leads to infer a tree (right) that differs from the true phylogeny (left)

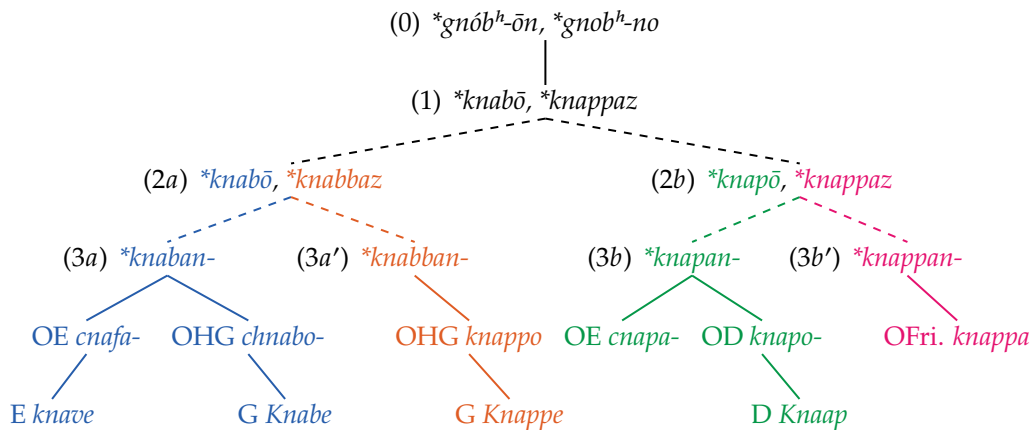


**Figure 7:** Incomplete lineage sorting from the polymorphism in the words for ‘fire’ in Indo-European

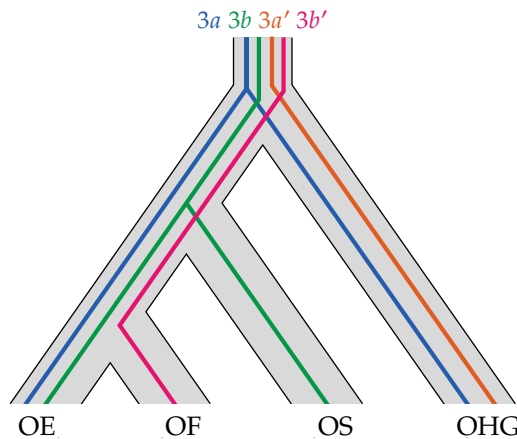
Another type of incomplete lineage sorting can happen by analogical levelling of morphological alternations, as in the example of the reflexes of Proto-Germanic *\*knabō*, *\*knappaz* ‘boy’ (Figure 8, from Jacques & List 2019, based on data from Kroonen 2011: 71, 128 and Kroonen 2013: 294). Unlike Figure 7, where only words with the same meaning are considered, in Figure 8 reflexes that have undergone semantic innovations in some languages are included, in particular the meaning ‘squire’. Applying the known sound laws backwards to the attested forms, at least four distinct proto-forms have to be reconstructed: *\*knaban-*, *\*knapan-*, *\*knabban-* and *\*knappan-*. Some languages even have more than one reflex of this etymon, with distinct and specialized meanings. Our detailed knowledge of the history of the Germanic languages allows us to go beyond merely stating the presence of irregular correspondences: doublets such as those in Figure 8 are due to the effect of Kluge’s law (the change from *\*-Cn-* to a geminate voiceless stop in pretonic position, where *\*C* represents a stop) on the endings of *n*-stem nouns in pre-Proto-Germanic (stage 0; Kluge 1884, Kroonen 2011). The paradigm of the noun ‘boy’, like all nouns belonging to the same declension class, was subject in Proto-Germanic (stage 1) to a morphological alternation between *\*-b-* and *\*-pp-*. This alternation was variously levelled as *\*-b- : \*-bb-* or *\*-p- : \*-pp-* by stage 2. However, within a single language, levelling did not apply in the same way to all items belonging to this declension class. Some languages have both innovative (Old English *cnapa* from *\*knapan-*) and archaic (Old English *cnafa* from *\*knaban-*) forms for the same etymon. After the loss of the *\*-b- : \*-pp-* alternation, all languages went through a second wave of analogy, generalizing either the stem of the nominative (archaic *\*knaban-* or innovative *\*knapan-*) or

that of the genitive (archaic *\*knappan-* or innovative *\*knabban-*), resulting in the four variants attested across Germanic languages. The distribution of these proto-forms conflicts with the accepted phylogeny of the Germanic languages. For instance, the proto-forms (3a) and (3b) group Old English with respectively Old High German and Old Dutch, to the exclusion of Old Frisian, the actual closest relative of English. To account for this discrepancy, Figure 9 maps the etymon tree (Fig. 8) onto the known phylogeny of the West Germanic languages.

The existence of incomplete lineage sorting implies that no single innovation can constitute sufficient evidence for a subgrouping hypothesis, whether it concerns the replacement of an etymon by an unrelated etymon for a particular meaning (Figure 7) or morphological innovations based on the same root form (Figures 8 and 9).



**Figure 8:** Incomplete lineage sorting in the levelling of *\*knabō, \*knappaz* ‘boy’ in Germanic (OE: Old English, OHG: Old High German, OD: Old Dutch, OFri.: Old Frisian, E: English, G: German, D: Dutch)



**Figure 9:** Etymon tree (color lines) vs language tree (grey hull)

### 3 Inferring and interpreting trees

#### 3.1 Identifying shared innovations

A phylogenetic tree represents the sequence of historical changes between an ancestor and its descendant languages. Inferring a tree thus necessitates to identify not simply similarities that are likely due to common descent, but shared innovations that are indicative of a common history and subgrouping. When the common ancestor of a group of languages, or an ancient language close to that ancestor, is attested, as in the case of Latin with regards to the Romance languages, it is usually easy to identify which features are symplesiomorphies and which are synapomorphies. However, such cases are extremely rare among the language families of the world, and inferring which common features are retentions and which are innovations is generally difficult. Although sound changes are often directional, many changes are bidirectional or have no clear directionality. Lexical items with unknown etymology are particularly problematic as it is usually impossible to determine on a purely linguistic basis which are ancient and which are more recent creations. This is one of the reasons why modern phylogenetic methods do not require innovations to be determined in advance, but identify them during the process of tree inference (Brower & Schuh 2021: 113–141).

Moreover, in historical linguistics as in biology, innovations and retentions are relative concepts: a given feature can constitute a shared innovation with respect to a recent common ancestor, but a retention with respect to a more remote proto-language. For instance, from the point of view of Proto-Germanic, the fact that the verb ‘bite’ and its whole paradigm are cognate in Gothic (infinitive *beitan*, preterit 3SG *bait*, preterit 3PL *bitun*), Old English (infinitive *bītan*, preterit 3SG *bāt*, preterit 3PL *biton*) or Old High German (infinitive *bīzan*, preterit 3SG *beizz*, preterit 3PL *bizzun*) is a retention, but from the point of view of Indo-European, it is a common Germanic innovation, since the cognate root *\*b<sup>h</sup>eyd-* originally meant ‘split’ as in Sanskrit *bhinatti* and Latin *findō*, and since it was originally a *\*-n-* infixing athematic present stem (*\*b<sup>h</sup>i<n(é)>d-*), whereas the Germanic present originates from a thematic conjugation *\*b<sup>h</sup>eyd-e/o-* (Seebold 1970), which spread and replaced nearly all athematic conjugations in Germanic.

#### 3.2 Methods of tree inference

Until the widespread adoption of modern quantitative methods, trees were traditionally inferred by hand (e.g. Leskien 1876, Brugmann 1884, Birnbaum & Puhvel 1966, Goddard 1994, Blust 1999, Georg 2004, Sagart 2004, Hill 2011, Pellard 2015, Campbell 2021: 230–242). Trees were constructed step by step on the basis of a carefully selected list of exclusively shared innovations, using an argumentative method similar to that of Hennig (1950, 1979) in biology. This procedure is problematic, not only because it poses the risk of subjectivity and bias in the selection of data (“cherry-picking”), but also because it requires innovations to be identified in advance, which is not obvious except in the case of changes with strong directionality. By contrast, computational phylogenetic methods usually avoid selection bias

by using a standard list of characters, do not need to identify innovations beforehand, and they allow the use of a much larger amount of data, both in terms of the number of characters and the number of languages compared (the number of possible rooted bifurcating trees is more than 34 million for 10 languages, 8 followed by 21 zeroes for 20 languages, and for more than 50 languages it approaches the number of atoms in the observable universe, far more what can be handled manually; [Felsenstein 2004: 24](#)). In any case, phylolinguistics methods should only be applied to languages known to be related: they cannot be used to demonstrate that languages belong to a single family; instead, they allow us to understand the internal relationships within a family.

A first type of quantitative method of tree inference is that of distance-based methods, such as neighbor-joining ([Saitou & Nei 1987](#)), or UPGMA ([Sokal & Michener 1958](#)), which are based on pairwise comparisons of languages. They are straightforward and fast: they simply count the number of dissimilarities and use these distances to build a tree ([Felsenstein 2004: 147–175](#), [Van de Peer & Salemi 2009](#)). However, distance-based methods fail to distinguish between shared innovations and shared retentions, and this loss of crucial information often leads to inferring misleading trees that tend, for example, to group conservative languages together. Moreover, observed distances do not reflect actual distances if changes occur at different rates in different branches. Lexicostatistics ([Sankoff 1973](#), [Embleton 1986](#), [Dyen, Kruskal & Black 1992](#)) is a classic example of such a method in linguistics.

The method of maximum parsimony, of which there are several variants, aims to find the tree with the smallest number of changes, i.e. common similarities are preferably interpreted as due to shared ancestry rather than parallel developments or reversals, and the tree that minimizes the number of ad hoc assumptions of such homoplasies is preferred ([Hennig 1950, 1979](#), [Felsenstein 2004](#), [Swofford & Sullivan 2009](#)). Maximum parsimony is less used nowadays, especially because of a problem known as *long-branch attraction* that causes rapidly evolving species or languages to be erroneously grouped together ([Felsenstein 1978](#)).

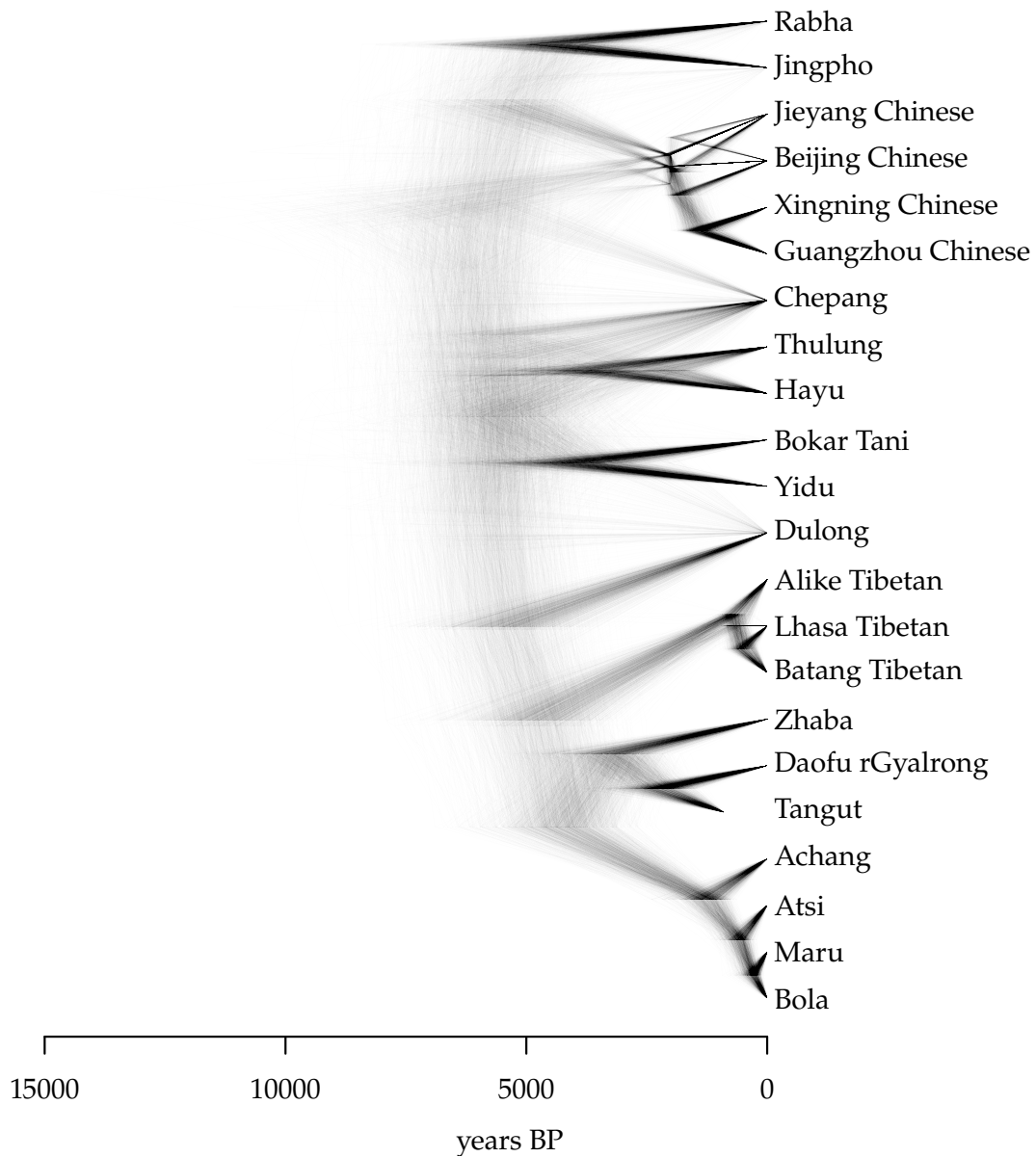
In phylogenetics, parsimony and distance-based methods are sub-optimal and tend to be biased or overconfident about deep relationships ([O'Reilly et al. 2016](#), [Puttick et al. 2017](#)). This observation is even more true for language phylogenies: whether because of limited amounts of data or because parts of the data did not actually evolve along a tree, deep relationships tend to be blurry, and the uncertainty in the reconstruction must therefore be made explicit. For this reason, methods based on explicit probabilistic models of evolution have been developed. The method of maximum likelihood searches for the single tree that maximizes the probability of the observed data, under a particular probabilistic model of evolution ([Felsenstein 2004: 248–274](#), [Schmidt & von Haeseler 2009](#)). Bayesian methods, on the other hand, search for a set of plausible trees given the data and the model. The result is a collection of trees, whose summary allows one to quantify the uncertainty of the inference ([Felsenstein 2004: 288–306](#), [Ronquist, van der Mark & Huelsenbeck 2009](#), [Drummond & Bouckaert 2015](#), [Hoffmann et al. 2021](#)). These trees typically agree on certain parts of the history, which are thus reconstructed with high confidence, but disagree on other parts, which remain uncertain.

The vast majority of quantitative works in the emerging field of phylo linguistics uses the Bayesian approach (Bowers 2018, Greenhill, Heggarty & Gray 2020, Hoffmann et al. 2021). Bayesian phylo linguistics have been applied to many families of languages, including Indo-European (Gray & Atkinson 2003, Ryder & Nicholls 2011, Chang et al. 2015, Heggarty et al. 2023), Uralic (Lehtinen et al. 2014), Bantu (Currie et al. 2013, Grollemund et al. 2015), Dravidian (Kolipakam et al. 2018), Sino-Tibetan (Sagart et al. 2019, Zhang et al. 2019), Austronesian (Gray, Drummond & Greenhill 2009), Timor-Alor-Pantar (Kaiping & Klammer 2022), Pama-Nyungan (Bowers & Atkinson 2012), Uto-Aztecan (Greenhill et al. 2023), Tupí-Guaraní (Gerardi & Reichert 2021), sign languages (Abner et al. 2024), and others. Bayesian methods rely on the Tree model and a probabilistic model of how languages change along a tree (WBCDL055 Computational phylogenetic models). The two models are distinct, the latter being manageable only thanks to the former. For example, a model of language change might describe how lexical items arise and disappear, but these models are mathematically and computationally tractable only if we assume that the history of languages can be modelled by a tree. Finally, any Bayesian inference procedure requires a *prior distribution*, which can describe all beliefs held about the tree and parameters before the data are taken into account. Practitioners should typically use uninformative priors (for example a prior which gives equal weight to all possible topologies, and a prior under which all reasonable root ages have non-negligible probability) so that the information in the posterior distribution is influenced by the data and not by the prior. Prior robustness checks, wherein the practitioner verifies that different prior distributions lead to similar posterior distributions, should also be used to ensure that the prior does not have an undue weight on the conclusion (Robert 2007).

### 3.3 Summarizing and interpreting the tree topology

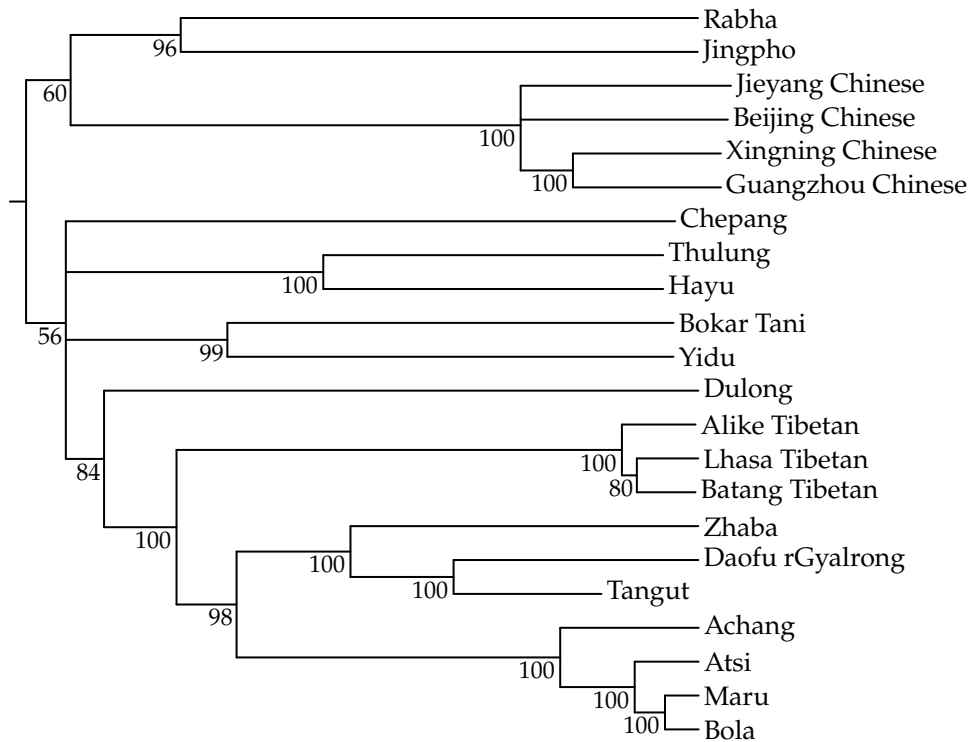
Tree inference can result in several or even many trees. This is particularly true for Bayesian methods, which yield a posterior sample of thousands of trees, but also for any method that uses statistical techniques to estimate the reliability of the results (Felsenstein 2004: 335–363). In order to estimate this uncertainty and interpret the results, the posterior sample thus needs to be summarized. We illustrate here different methods with a posterior sample of 2 000 trees of 21 Sino-Tibetan languages subsampled from Sagart et al. (2019).

A densitree (Bouckaert 2010) is a useful visualization method in which all the resulting trees of the analysis are overlaid and displayed semi-transparently (Figure 10). This directly conveys the (un)certainty of both the tree topology and the internal node ages, but the deeper parts of the tree are often difficult to read. For example, Figure 10 shows that Achang, Atsi, Maru and Bola form a subgroup with high probability; the same is true of Common Chinese (comprising the four different Chinese varieties), even though its internal structure is only partly resolved probably due to undetected borrowings from Classical Chinese. On the other hand, the position of Chepang is highly uncertain, as is the deeper part of the tree.



**Figure 10:** Densitree of Sino-Tibetan languages

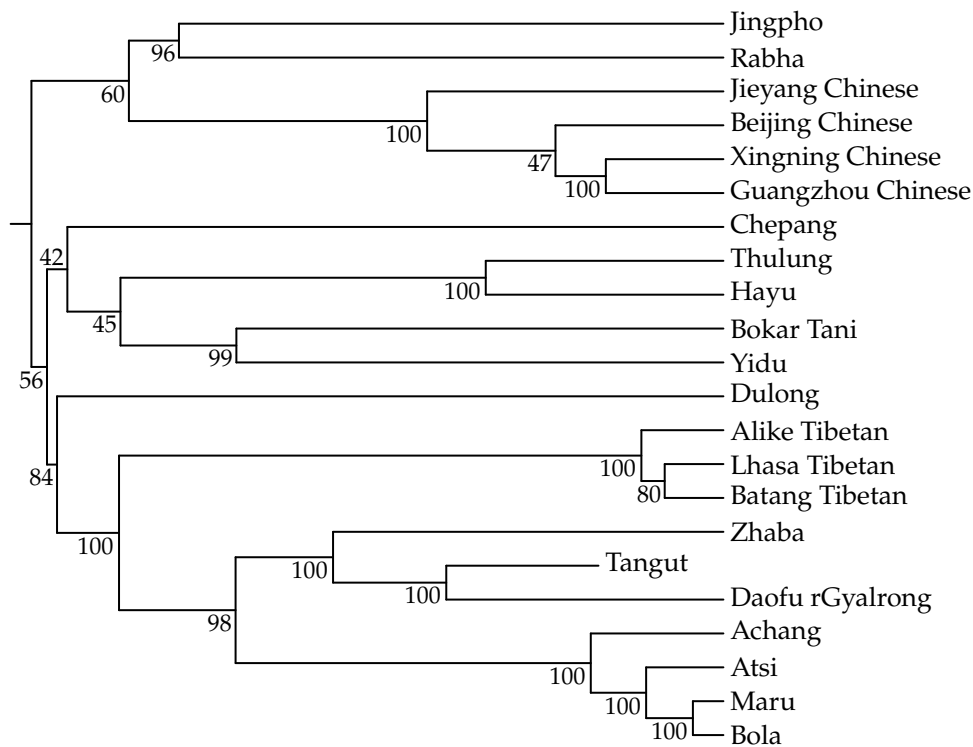
A consensus tree (Figure 11) is a single tree that only displays subgroups for which a number of trees in the sample agree. In a majority-rule consensus (MRC) tree, a node is displayed if it exists in more than 50% of the sampled trees. The probability of each node is displayed (but may be omitted for nodes with support close to 100%). For certain aspects of the topology, the uncertainty is too large and no bifurcation has support above 50%, in which case the consensus tree is multifurcating rather than bifurcating. It is not uncommon



**Figure 11:** Majority-rule consensus tree of Sino-Tibetan languages

to observe a single node from which many branches stem (a “rake” or “star”), indicating that this part of the topology is highly uncertain. In the majority-rule consensus tree of Figure 11, the probability that the four Chinese varieties form a subgroup is 100%, i.e. this subgroup exists in all trees in the sample, but its internal structure is highly uncertain and is displayed as multifurcating, since no bifurcating topology has a probability above 50%. The uncertainty about the deeper part of the topology observed on the densitree of Figure 10 is here evident from the multifurcation leading to Chepang. The probability that Chinese and Sal (Rabha and Jingpho) form a clade is 60%, as indicated on the consensus tree, but this does not mean that the probability that Chinese-Sal is a primary branch is 60% too since that probability also includes cases where the Chinese-Sal clade is nested within another clade. The highest probability (32%) is actually that Chinese is a primary branch, with all other languages forming a clade, but that probability is too low to appear on the consensus tree, which instead shows the more probable Chinese-Sal clade. The probability that the first bifurcation is Chinese-Sal vs. all other languages is actually only 19%, but this cannot be read from the consensus tree. In any case, no firm conclusion can be drawn about the first bifurcation of the Sino-Tibetan tree.

A maximum clade credibility (MCC) tree is the “best” binary tree that summarizes the sample, defined as the tree topology that maximizes the product of posterior support



**Figure 12:** Maximum clade credibility tree of Sino-Tibetan languages

for each node. The corresponding probability is displayed at each node (Figure 12). This tends to obscure uncertainty and possibly give a false sense of security since it is more difficult to detect limitations of the Tree model with an MCC than with a consensus tree or a densitree. An MCC often includes poorly supported clades, especially for data including relatively few characters (O'Reilly & Donoghue 2018), which is usually the case in linguistics. Similarly, a maximum *a posteriori* (MAP) tree is simply the tree with the greatest overall posterior probability (for both topology and branch lengths) in the sample, but this hides the uncertainty in the results and discards most of the information contained in the output. We advise against the temptation to summarize a posterior distribution as a single binary tree (such as the MCC or MAP tree). We advocate for summaries which are candid about the topological uncertainty which often arises in linguistic phylogenies, such as densitrees and consensus trees.

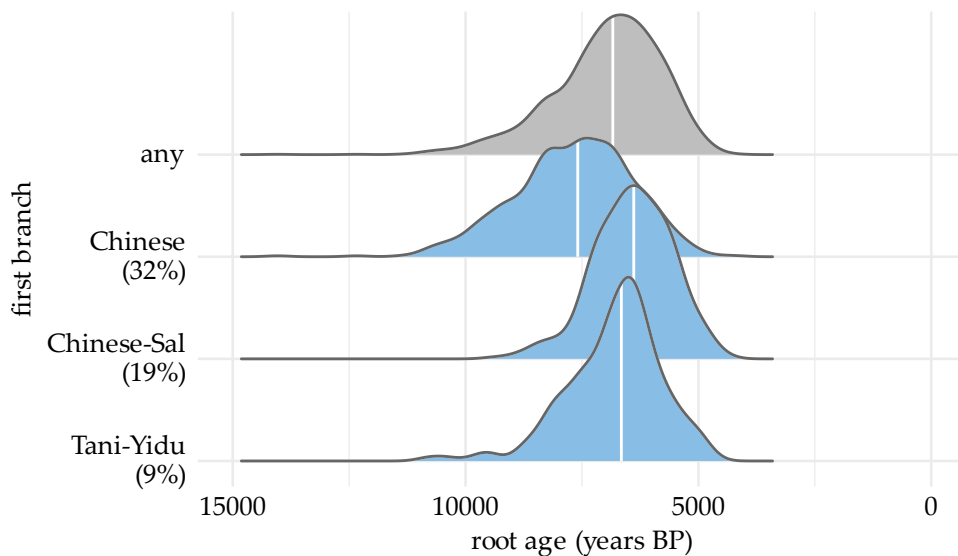
### 3.4 Summarizing and interpreting the dates

The output of a Bayesian phylolinguistics study is a distribution representing the uncertainty not only of the topology but also of the different ages and evolutionary parameters, such as rates of change. These can be simply summarized with histograms or density plots showing

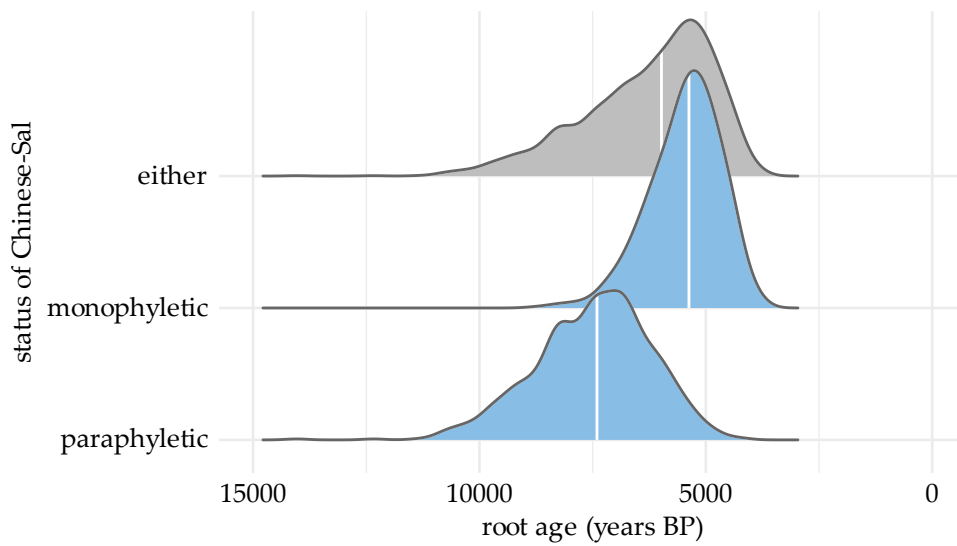
the range of plausible values. Note that it is possible for the topology to be well reconstructed, i.e. there is low uncertainty, but for the root age to be highly uncertain. Conversely, it is possible that the topology is uncertain but not the root age, e.g. because there are many different branching scenarios although all agree on a similar root age.

Figure 13 gives density probability estimates for the root age in the Sino-Tibetan example, for all trees and several subsets restricted to those trees that contain one of the three most probable first branches, e.g. with Common Chinese vs. all other languages being the first bifurcation. There are slight differences in the inferred root age depending on these topology configurations, which translates into more uncertainty in the overall distribution. Note that attempting to read the age of an internal node directly on a consensus, MCC, or MAP tree can be misleading, and estimated ages should always be computed from the posterior sample.

The distribution of the age of internal nodes can also be displayed. When dealing with such internal node ages, it should be borne in mind that the very existence of internal nodes can be uncertain. In the Sino-Tibetan example, it is easy to consider the age of the most common recent ancestor (MRCA) to Achang, Atsi, Maru and Bola, since these languages form a clade with probability 100%. If, on the other hand, we want to consider the age of the MRCA to Rabha, Jingpho and the four Chinese varieties (Chinese-Sal), we must bear in mind that these languages do not necessarily form a clade: their MRCA can also have other descendants. The distribution of the age of this MRCA is shown in Figure 14, depending on whether these languages form a clade or not. When Chinese and Sal belong to separate parts of the tree, their MRCA tends to be older, which is also reflected in the overall distribution that expresses the full range of possible values for the age of the MRCA.



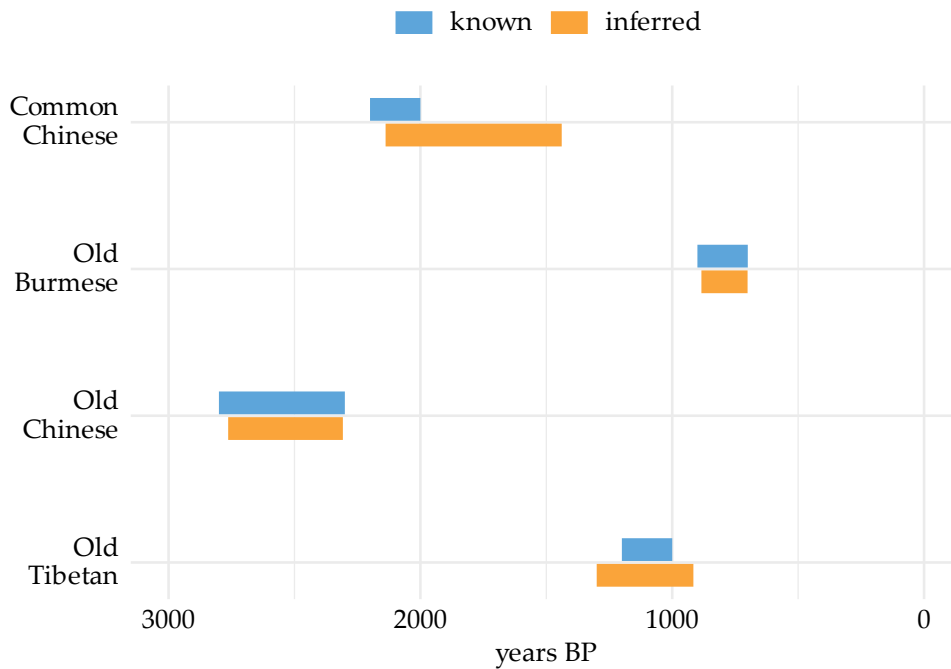
**Figure 13:** Probability densities for Sino-Tibetan root age, with the median indicated by a white line, for all trees and for subsets restricted to trees containing a certain first branch; percentages indicate the probability that the given subgroup constitutes a first branch



**Figure 14:** Probability densities of the age of the MRCA to Common Chinese and Sal (Rabha and Jingpho), with the median indicated by a white line, for all trees and depending on whether these languages form a clade (monophyletic) or not (paraphyletic)

The first attempts at dating ancient languages using quantitative methods gave birth to glottochronology (Swadesh 1952), which assumes that basic vocabulary evolves at a constant rate through time, space and meanings. Bergsland & Vogt (1962) famously showed that the assumptions of glottochronology do not hold and that the inferred ages are not reliable, and quantitative methods for dating languages subsequently fell into disrepute (McMahon & McMahon 2006). It can be argued that modern methods of estimating dates based on phylogenies qualify as glottochronology, and that the term should be redefined accordingly. In any case, modern methods do not suffer from the shortcomings of traditional glottochronology since the models used are more elaborate and robust, and they do not assume a universal rate of change. They also estimate and make explicit the uncertainty in their results. In some cases, this leads to answering “we don’t know” to questions of dating, but otherwise it shows the confidence we can have in the inference made.

Furthermore, the validity of the dates inferred is demonstrated by validation methods which show that known ages are well reconstructed. Phylolinguistic analyses typically rely on a few known ages: either ancient languages whose ages we know, up to some error, or internal node ages corresponding to a split which we can date from independent evidence. In the Sino-Tibetan example, four ages were known, those of the ancient languages Old Chinese, Old Tibetan, and Old Burmese, and that of the split of Common Chinese. These known time constraints are used in the inference procedure to help infer the parameters, but can also be used to validate the method. We can pretend to ignore one of the ages, estimate the age of the corresponding node, and then check that the estimated age matches the known age. Ryder & Nicholls (2011) provide a formal statistical test of the fit, but in many cases a



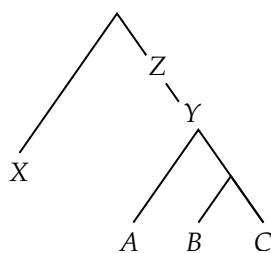
**Figure 15:** Cross-validation of known ancient ages in Sino-Tibetan; bars for the inferred ages represent the 95% highest posterior density intervals

simple visual representation such as Figure 15 may suffice. The inferred ages naturally often show more uncertainty, but they largely overlap with the known ages. Note that internal node ages are typically reconstructed more robustly than ancient leaf ages (Ryder 2010).

### 3.5 The status of ancient and ancestral languages

Ancient languages that are recorded in writing and can be dated, such as Akkadian, Latin or Old Chinese, are invaluable sources of information for historical linguistics, including for inferring phylogenies. They can be used to determine which features are innovations and retentions, and to calibrate the rate of change along a phylogenetic tree. This can be then used to evaluate the ages of all the nodes in the tree. However, ancient is not synonymous with ancestral (Figure 16), and ancient languages are probably never the exact ancestors of later languages. Even in seemingly obvious cases, such as Latin and the Romance languages (Meillet 1925: 7–11, Garnier 2016), or Old Tibetan and the modern Tibetic languages (Sangsrgyas Tshering 2020), we find some lexical and morphological features that cannot be derived from the corresponding ancient written language and must derive from close but distinct unwritten varieties.

For instance, Garnier (2012) shows that the Romance proto-form *\*rōkka-* f. ‘rock’ (reflected by Italian *rocca* and French *roche*) is not inherited from Latin *rūpēs* f. ‘rock’, but rather that both originate from a Proto-Italic etymon *\*rowp-í-* ‘cliff’ (plural *\*rowp-éyes* ‘the rocks’), from



**Figure 16:** Ancient and ancestral languages: *A*, *B* and *C* are observed extant languages; *X* is an ancient language, which might exist in the record, but from which *A*, *B* and *C* do not descend; *Y* and *Z* are ancestral languages, *Y* being the most recent common ancestor to languages *A*, *B* and *C*

the root *\*rewp* ‘break’ (Rix et al. 2001: 510–511). The expected Latin form *†rūps* for the nominative singular (from *\*rowpís*) was renewed, and only the genitive plural *rūpium* (from *\*rowpi-yōm*) is inherited from Proto-Italic. By contrast, Proto-Romance *\*rōkka-* ‘rock’ arose from the alternating paradigm *\*rōpika* NOM.SG : *\*rōkkārum* GEN.PL, which arose as the result of a sound law whereby unstressed short vowels between two strong syllables are deleted: *\*rōkkārum* < *\*rōpkārum* < *rōpikārum*. The etymon *\*rōpika* itself is derived from the diminutive *\*rowpikelā* from Proto-Italic *\*rowpí-*, with a distinct treatment of *\*ow* as *ō*, reflecting a plebeian sociolect. While most Proto-Romance etyma can be derived from Latin by regular sound correspondences, forms originating from etyma non-attested in Latin like *\*rōkka-* ‘rock’ are not rare, and Classical Latin is thus not the exact ancestor of Romance languages, even though it is close to it.

### 3.6 Evaluating the adequacy of the Tree model

The Tree model is a simplification of reality (§1.2), and like for any other model, this simplification is reasonable and useful in certain cases, but unreasonable or useless in others. The collection of trees returned by a Bayesian analysis already gives an overview of the extent to which the data can be explained by a tree. When there are blurry parts, as near the root of the tree in Figure 10, it is however difficult to assess whether this is due to data that do not follow a tree, or simply to data that are not informative enough of certain parts of the history.

Several measures of treeness have been developed to estimate how well the Tree model fits the data, each with its benefits and flaws (Münkemüller et al. 2012). Popular measures include Pagel’s  $\lambda$  (Pagel 1999), Moran’s  $I$  (Gittleman & Kot 1990), and Bloomberg’s  $K$  (Blomberg, Garland & Ives 2003), which is described in the context of linguistics by Macklin-Cordes, Bowern & Round (2021). For instance, Pagel’s  $\lambda$  is used to measure the treeness of continuous data, i.e. “whether the shared evolutionary histories as specified by the phylogeny produce the patterns of similarity observed in the data” (Pagel 1999: 883). Computed from the total lengths of shared branches between all pairs of tips in a tree and a Brownian motion model of trait evolution,  $\lambda$  takes values between 0 and 1, with a value of  $\lambda = 1$  indicating that

the phylogeny perfectly explains all the variation in trait values, and a value of  $\lambda = 0$  that none of the variation is explained by the phylogeny. The TIGER score has also been proposed as a measure of the internal consistency of a dataset and its treeness prior to tree inference (Syrjänen et al. 2021, List 2022). A more complete strategy to evaluate treeness has recently become available with models and methods that propose both a means of testing the fit of the Tree model, and an alternative representation for when the fit is poor (Kelly & Nicholls 2017, Neureiter et al. 2022). The question of whether a tree is an adequate representation of how the data arose comes on top of questions of the fit of any specific model of evolutionary change (Carstens et al. 2022).

## 4 Beyond classification

The Tree model and phylogenetic methods have been successfully applied not only to the study of the subgrouping of many language families, but also to the study of a wide range of questions beyond classification, especially about the evolution of languages and human prehistory.

### 4.1 Studying language evolution with trees

Determining which features are innovations and which are retentions is fundamentally a problem of phylogeny (Jacques & Pellard 2021), and model-based phylogenies allow for *ancestral state reconstruction* (Pagel, Meade & Barker 2004). Given a phylogenetic tree (or collection of trees) and a probabilistic model of linguistic change, it is possible to reconstruct the features of a proto-language, e.g. the morphosyntactic alignment and other grammatical features of Proto-Indo-European (Carling & Cathcart 2021a,b), argument structure constructions and their case frames in Proto-Germanic (Dunn et al. 2017), as well as alignment (Phillips & Bower 2022), numeral systems (Zhou & Bower 2015), and color term systems (Haynie & Bower 2016) in Pama-Nyungan. The reconstruction of proto-forms can also be attempted with such methods (Bouchard-Côté et al. 2013).

It is also possible to study trends in the evolution of linguistic features, such as their relative stability and rate of change (Pagel, Atkinson & Meade 2007, Greenhill et al. 2017), or correlations in their evolution (Dunn et al. 2011, Goldstein 2022, Round, Dockum & Ryder 2022), as well as rates and tempo of language diversification (Atkinson et al. 2008).

Phylogenetic methods can also solve a major problem in linguistic typology: that of phylogenetic correlation and sampling. Observations in different languages cannot be treated as independent when looking for statistical tendencies, since similarities may be due to common ancestry rather than typological tendencies, as closely related languages tend to resemble each other (the so-called “Galton’s problem”). Phylogenetic methods can solve for this issue by taking into account the topology and branch lengths of the phylogenetic tree of the languages under study (Macklin-Cordes & Round 2022, Bromham 2024).

## 4.2 Linguistic phylogenies and human prehistory

Linguistic phylogenies reflect the history of their speakers, and it is interesting to evaluate the concordance of genetic, linguistic and cultural history by comparing their different trees (Matsumae et al. 2021, Barbieri et al. 2022). Though we can expect that speakers transmit their genes, language, and culture to their descendants, mismatches between their respective histories are at least as interesting as matches since they are indicative of historical events worthy of further investigation.

Language phylogenetic trees, especially dated ones, are particularly important for determining the geographical origin (*Urheimat*) and dispersal routes of a language family, as well as for potentially identifying language families with archaeological cultures and ancient populations (Gray & Atkinson 2003, Gray, Drummond & Greenhill 2009, Chang et al. 2015, Sagart et al. 2019, Lazaridis et al. 2022, Heggarty et al. 2023). The traditional view is that the homeland of a language family is most probably located near the area of its maximal linguistic diversity (Sapir 1921), because languages located close to the homeland have more time to diverge from each other, while those located farther away may result from a limited number of migrations and bottlenecks effects, and have a greater chance of all belonging to a single subbranch. The diversity in question here is not the number of distinct languages per areal unit, but rather the number of different tree branches and the depth of the divergence between the languages in a given area. The maximal diversity principle has proved valid in some cases, e.g. Austronesian and its maximally diverse homeland located in Taiwan (Blust 1988), but not in all, especially since there are cases where a former diversity has been obliterated afterwards. For instance, both linguistic palaeontology and archaeology suggest that Sino-Tibetan originates from the Yellow River, although the current linguistic diversity of Northern China is very low (Sagart et al. 2019).

More recently, the methods of phylogeography (Avice 2009, Lemey et al. 2010) have been applied to linguistic data in order to infer geographical origin and dispersal routes (e.g. Bouckaert et al. 2012, Koile et al. 2022), though the adequacy of these methods is still being investigated (Wichmann & Rama 2021, Neureiter et al. 2021).

## 5 Conclusions

The Tree model has proved to be a useful representation of the history and relationships of languages, and an invaluable tool for studying language evolution and linking the history of languages to the history of their speakers. This has allowed some previously unresolved questions to be answered while leaving other questions open, and validation tests show that the answers are reliable. Phylolinguistics has not only been successful but has also made considerable progress in recent decades, and further improvements are being developed. The tree models used today are more sophisticated and flexible than in the past.

Refinements we hope to see developed in the future include: better models for morphosyntactic features, and joint models for lexical, phonological and morphosyntactic data; enrichments of trees to include borrowings and diffusion processes, and better interaction

with geographical and environmental information; comparison of languages with other cultural phylogenies; using dated phylogenies from other families to calibrate dates in language families lacking ancient languages. Further developments in computer-assisted language comparison and cognate identification (List, Greenhill & Gray 2017, List 2019, Clarté & Ryder 2022) could also help overcome the need for manual annotation of cognates by experts and allow to apply phylogenetic methods to lesser-studied language families whose historical phonology is not yet well understood.

Nevertheless, the Tree model remains a simplification, and even the most ardent advocates of trees would not claim that a single tree can account for all the complexities and details of a language's evolution. It suffers from limitations and relies on assumptions that do not necessarily always hold. In particular, the Tree model and phylogenetic methods assume that languages are related, and they are therefore unable to demonstrate relatedness. It is therefore not advisable to apply these methods to controversial language families (Tian et al. 2022): the result will by definition be a tree, but this is of little value since there is no guarantee that it represents a true phylogeny. In any case, the results can only be good as the input, and no amount of fancy mathematics can compensate for bad data.

Still, the Tree model remains the best tool at our disposal for now, and refined models and methods are being developed in order to overcome current limitations. More and more linguists are now adopting phylogenetic methods, breaking away from the tradition of a priori skepticism and pessimism towards computational and statistical methods in historical linguistics (Sankoff 1973: 96, Atkinson & Gray 2006, Verkerk 2017). We can hope this will further advance our understanding of language evolution and help bring together evidence from language, culture, and genetics to provide a more complete picture of the human past. In this sense, if an imperfect and inaccurate model is a fiction, it nevertheless may be more useful and virtuous than a more realistic and complex but less constrained model that fails to make precise predictions, except that anything goes.

In this sense, if an imperfect and inaccurate model is a fiction, it nevertheless may be more useful and virtuous than a more realistic and complex but less constrained model that fails to make precise predictions, except that anything goes.

## Code and data

The R code and data to reproduce the figures in this chapter are available from <https://doi.org/10.5281/zenodo.11131508>.

## Related articles

WBCDL036 Contact and borrowing; WBCDL043 Mixed languages; WBCDL055 Computational phylogenetic models; WBCDL059 Subgrouping; WBCDL060 Wave model: Diffusion, convergence, and differentiation

## Acknowledgments

We would like to express our sincere thanks to the editor, Joe Salmons, an anonymous reviewer, Alexandre François, Siva Kalyan, Erich Round, Simon Greenhill, and all the Academia session participants for their valuable contributions, corrections and constructive insights that helped enhance the quality of this chapter.

## References

- Abner, Natasha, Grégoire Clarté, Carlo Geraci, Robin J. Ryder, Justine Mertz, Anah Salgat & Shi Yu. 2024. Computational phylogenetics reveal histories of sign languages. *Science* 383(6682). 519–523. <https://doi.org/10.1126/science.add7766>.
- Aikio, Ante. 2007. Etymological nativization of loanwords: A case study of Saami and Finnish. In Ida Toivonen & Diane Nelson (eds.), *Saami linguistics*, 17–52. Amsterdam: John Benjamins. <https://doi.org/10.1075/cilt.288.03aik>.
- Anttila, Raimo. 1989. *Historical and comparative linguistics*. Amsterdam: John Benjamins. <https://doi.org/10.1075/cilt.6>.
- Atkinson, Quentin D. & Russell D. Gray. 2006. How old is the Indo-European language family?: Progress or more moths to the flame? In Peter Forster & Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, 91–109. Cambridge: McDonald Institute for Archaeological Research.
- Atkinson, Quentin D., Andrew Meade, Chris Venditti, Simon J. Greenhill & Mark Pagel. 2008. Languages evolve in punctuational bursts. *Science* 319(5863). 588. <https://doi.org/10.1126/science.1149683>.
- Auderset, Sandra, Simon J. Greenhill, Christian T. DiCano & Eric W. Campbell. 2023. Subgrouping in a ‘dialect continuum’: A Bayesian phylogenetic analysis of the Mixtecan language family. *Journal of Language Evolution* 8(1). 33–63. <https://doi.org/10.1093/jole/lzad004>.
- Avice, John C. 2009. Phylogeography: Retrospect and prospect. *Journal of Biogeography* 36(1). 3–15. <https://doi.org/10.1111/j.1365-2699.2008.02032.x>.
- Bakker, Peter. 2017. Typology of mixed languages. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *The Cambridge handbook of linguistic typology*, 217–253. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316135716.008>.
- Barbançon, François, Steven N. Evans, Luay Nakhleh, Don Ringe & Tandy Warnow. 2013. An experimental study comparing linguistic phylogenetic reconstruction methods. *Diachronica* 30(2). 143–170. <https://doi.org/10.1075/dia.30.2.01bar>.
- Barbieri, Chiara, Damián E. Blasi, Epifanía Arango-Isaza, Alexandros G. Sotiropoulos, Harald Hammarström, Søren Wichmann, Simon J. Greenhill, Russell D. Gray, Robert Forkel, Balthasar Bickel & Kentaro K. Shimizu. 2022. A global analysis of matches and mismatches between human genetic and linguistic histories. *Proceedings of the National Academy of Sciences* 119(47), e2122084119. <https://doi.org/10.1073/pnas.2122084119>.
- Baum, David A. & Stacey D. Smith. 2013. *Tree thinking: An introduction to phylogenetic biology*. Greenwood Village: Roberts.
- Bergsland, Knut & Hans Vogt. 1962. On the validity of glottochronology. *Current Anthropology* 3(2). 115–153. <https://doi.org/10.1086/200264>.

- Birnbaum, Henrik & Jaan Puhvel (eds.). 1966. *Ancient Indo-European dialects: Proceedings of the conference on Indo-European linguistics held at the University of California, Los Angeles, April 25–27, 1963*. Berkeley: University of California Press.
- Blomberg, Simon P., Theodore Garland & Anthony R. Ives. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* 57(4). 717–745. <https://doi.org/10.1111/j.0014-3820.2003.tb00285.x>.
- Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.
- Blust, Robert. 1988. The Austronesian homeland: A linguistic perspective. *Asian Perspectives* 26(1). 45–67. <http://hdl.handle.net/10125/16918>.
- Blust, Robert. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun & Paul Jen-kuei Li (eds.), *Selected papers from the eighth international conference on austronesian linguistics*, 1–54. Taipei: Academia Sinica.
- Blust, Robert & Stephen Trussel. 2010. *Austronesian comparative dictionary*. <http://www.trussel2.com/acd/>.
- Bonfante, Giuliano. 1946. “Indo-Hittite” and areal linguistics. *The American Journal of Philology* 67(4). 289–310. <https://doi.org/10.2307/290681>.
- Bouchard-Côté, Alexandre, David Hall, Thomas L. Griffiths & Dan Klein. 2013. Automated reconstruction of ancient languages using probabilistic models of sound change. *Proceedings of the National Academy of Sciences* 110(11). 4224–4229. <https://doi.org/10.1073/pnas.1204678110>.
- Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard & Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097). 957–960. <https://doi.org/10.1126/science.1219669>.
- Bouckaert, Remco R. 2010. DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics* 26(10). 1372–1373. <https://doi.org/10.1093/bioinformatics/btq110>.
- Bowern, Claire. 2018. Computational phylogenetics. *Annual Review of Linguistics* 4(1). 281–296. <https://doi.org/10.1146/annurev-linguistics-011516-034142>.
- Bowern, Claire & Quentin Atkinson. 2012. Computational phylogenetics and the internal structure of Pama-Nyungan. *Language* 88(4). 817–845. <https://doi.org/10.1353/lan.2012.0081>.
- Bowern, Claire, Patience Epps, Russell Gray, Jane Hill, Keith Hunley, Patrick McConvell & Jason Zentz. 2011. Does lateral transmission obscure inheritance in hunter-gatherer languages? *PLOS One* 6(9), e25195. <https://doi.org/10.1371/journal.pone.0025195>.
- Box, George E. P. 1979. Robustness in the strategy of scientific model building. In Robert J. Launer & Graham N. Wilkinson (eds.), *Robustness in statistics*, 201–236. New York: Academic Press. <https://doi.org/10.1016/B978-0-12-438150-6.50018-2>.
- Britain, David. 2013. Space, diffusion and mobility. In J. K. Chambers & Natalie Schilling (eds.), *The handbook of language variation and change*, 2nd edn., 469–500. Chichester: Wiley. <https://doi.org/10.1002/9781118335598.ch22>.
- Bromham, Lindell. 2024. Solving Galton’s problem: Practical solutions for analysing language diversity and evolution. *PsyArXiv*. <https://doi.org/10.31234/osf.io/c8v9r>.
- Brower, Andrew V. Z. & Randall T. Schuh. 2021. *Biological systematics: Principles and applications*. 3rd edn. Ithaca: Comstock. <https://doi.org/10.7591/cornell/9781501752773.001.0001>.
- Brugmann, Karl. 1884. Zur Frage nach den Verwandtschaftsverhältnissen der indogermanischen Sprachen. *Internationale Zeitschrift für Allgemeine Sprachwissenschaft* 1. 226–256. <https://archive.org/details/internationalez00unkngoog>.

- Bryant, David & Vincent Moulton. 2004. Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21(2). 255–265. <https://doi.org/10.1093/molbev/msh018>.
- Buck, Carl Darling. 1949. *A dictionary of selected synonyms in the principal Indo-European languages*. Chicago: University of Chicago Press.
- Campbell, Lyle. 2021. *Historical linguistics: An introduction*. 4th edn. Edinburgh: Edinburgh University Press. <https://doi.org/10.1515/9781474463133>.
- Campbell, Lyle & William J. Poser. 2008. *Language classification: History and method*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09780511486906>.
- Carling, Gerd & Chundra Cathcart. 2021a. Evolutionary dynamics of Indo-European alignment patterns. *Diachronica* 38(3). 358–412. <https://doi.org/10.1075/dia.19043.car>.
- Carling, Gerd & Chundra Cathcart. 2021b. Reconstructing the evolution of Indo-European grammar. *Language* 97(3). <https://doi.org/10.1353/lan.0.0253>.
- Carling, Gerd, Chundra Cathcart & Erich Round. 2021. Reconstructing the origins of language families and variation. In Nathalie Gontier, Andy Lock & Chris Sinha (eds.), *The Oxford handbook of human symbolic evolution*, 811–838. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198813781.013.34>.
- Carstens, Bryan C., Megan L. Smith, Drew J. Duckett, Emanuel M. Fonseca & M. Tereza C. Thomé. 2022. Assessing model adequacy leads to more robust phylogeographic inference. *Trends in Ecology & Evolution* 37(5). 402–410. <https://doi.org/10.1016/j.tree.2021.12.007>.
- Chang, Will, Chundra Cathcart, David Hall & Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91(1). 194–244. <https://doi.org/10.1353/lan.2015.0005>.
- Clarté, Grégoire & Robin J. Ryder. 2022. A phylogenetic model of the evolution of discrete matrices for the joint inference of lexical and phonological language histories. *ArXiv*. <https://doi.org/10.48550/arXiv.2206.12473>.
- Currie, Thomas E., Andrew Meade, Myrtille Guillon & Ruth Mace. 2013. Cultural phylogeography of the Bantu languages of sub-Saharan Africa. *Proceedings of the Royal Society B: Biological Sciences* 280(1762), 20130695. <https://doi.org/10.1098/rspb.2013.0695>.
- Dixon, R. M. W. & Alexandra Y. Aikhenvald (eds.). 2001. *Areal diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198299813.001.0001>.
- Drummond, Alexei J. & Remco R. Bouckaert. 2015. *Bayesian evolutionary analysis with BEAST*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09781139095112>.
- Dunn, Michael, Tonya Kim Dewey, Carlee Arnett, Thórhallur Eythórsson & Jóhanna Barðdal. 2017. Dative sickness: A phylogenetic analysis of argument structure evolution in Germanic. *Language* 93(1). e1–e22. <https://doi.org/10.1353/lan.2017.0012>.
- Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473(7345). 79–82. <https://doi.org/10.1038/nature09923>.
- Dyen, Isidore, Joseph B. Kruskal & Paul Black. 1992. An Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society* 82(5). 1–132. <https://doi.org/10.2307/1006517>.
- Elgin, Catherine Z. 2017. *True enough*. Cambridge: MIT press.
- Embleton, Sheila M. 1986. *Statistics in historical linguistics*. Bochum: Brockmeyer.

- Evans, Cara L., Simon J. Greenhill, Joseph Watts, Johann-Mattis List, Carlos A. Botero, Russell D. Gray & Kathryn R. Kirby. 2021. The uses and abuses of tree thinking in cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376(1828), 20200056. <https://doi.org/10.1098/rstb.2020.0056>.
- Felsenstein, Joseph. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Biology* 27(4). 401–410. <https://doi.org/10.1093/sysbio/27.4.401>.
- Felsenstein, Joseph. 1988. Phylogenies from molecular sequences: Inference and reliability. *Annual Review of Genetics* 22(1). 521–565. <https://doi.org/10.1146/annurev.ge.22.120188.002513>.
- Felsenstein, Joseph. 2004. *Inferring phylogenies*. Sunderland: Sinauer.
- Forster, Peter & Colin Renfrew (eds.). 2006. *Phylogenetic methods and the prehistory of languages*. Cambridge: McDonald Institute for Archaeological Research.
- Fox, Anthony. 1995. *Linguistic reconstruction: An introduction to theory and method*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198700005.001.0001>.
- François, Alexandre. 2015. Trees, waves and linkages: Models of language diversification. In Claire Bowerman & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 161–189. Abingdon: Routledge. <https://doi.org/10.4324/9781315794013>.
- Franklin-Hall, L. R. 2010. Trashing life's tree. *Biology & Philosophy* 25(4). 689–709. <https://doi.org/10.1007/s10539-010-9219-1>.
- Galtier, Nicolas & Vincent Daubin. 2008. Dealing with incongruence in phylogenomic analyses. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363(1512). 4023–4029. <https://doi.org/10.1098/rstb.2008.0144>.
- Garnier, Romain. 2012. Allomorphisme et lois de limitation rythmique en latin. *Bulletin de la Société de Linguistique de Paris* 107(1). 235–259. <https://doi.org/10.2143/BSL.107.1.2959705>.
- Garnier, Romain. 2016. Ouverture: Protoroman, latin et indo-européen. In Éva Buchi & Wolfgang Schweickard (eds.), *Dictionnaire étymologique roman (DÉRom)*. Vol. 2: *Pratique lexicographique et réflexions théoriques*, 2 vols., 219–256. Berlin: De Gruyter. <https://doi.org/10.1515/9783110453614-010>.
- Gąsiorowski, Piotr. 1999. The tree of language: A cladistic look at the genetic classification of languages. *Dialectologia et Geolinguistica* 1999(7). 39–58. <https://doi.org/10.1515/dig.1999.1999.7.39>.
- Geisler, Hans & Johann-Mattis List. 2013. Do languages grow on trees?: The tree metaphor in the history of linguistics. In Heiner Fangerau, Hans Geisler, Thorsten Halling & William Martin (eds.), *Classification and evolution in biology, linguistics and the history of science*, 111–124. Stuttgart: Steiner.
- Geisler, Hans & Johann-Mattis List. 2022. Of word families and language trees: New and old metaphors in studies on language history. *Moderna* 24(1–2). 135–148.
- Georg, Stefan. 2004. Unreclassifying Tungusic. In Carsten Naehrer (ed.), *Proceedings of the First International Conference on Manchu-Tungus Studies (Bonn, August 28–September 1, 2000)*. Vol. 2: *Trends in Tungusic and Siberian linguistics*, 2 vols., 45–57. Wiesbaden: Harrassowitz.
- Gerardi, Fabrício Ferraz & Stanislav Reichert. 2021. The Tupí-Guaraní language family: A phylogenetic classification. *Diachronica* 38(2). 151–188. <https://doi.org/10.1075/dia.18032.fer>.
- Gittleman, John L. & Mark Kot. 1990. Adaptation: Statistics and a null model for estimating phylogenetic effects. *Systematic Biology* 39(3). 227–241. <https://doi.org/10.2307/2992183>.
- Gleason, Henry A., Jr. 1961. Genetic relationship among languages. In Roman Jakobson (ed.), *Structure of language and its mathematical aspects: Proceedings of the twelfth Symposium in Applied Mathematics*,

- held in New York City, April 14–15, 1960, 179–189. Providence: American Mathematical Society. <https://doi.org/10.1090/psapm/012>.
- Goddard, Ives. 1994. The West-to-East cline in Algonquian dialectology. In William Cowan (ed.), *Actes du vingt-cinquième congrès des algonquinistes*, 187–211. Ottawa: Carleton University. <https://ojs.library.carleton.ca/index.php/ALGQP/article/view/616>.
- Goldstein, David. 2022. Correlated grammaticalization: The rise of articles in Indo-European. *Diachronica* 39(5). 658–706. <https://doi.org/10.1075/dia.20033.gol>.
- Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426(6965). 435–439. <https://doi.org/10.1038/nature02029>.
- Gray, Russell D., David Bryant & Simon J. Greenhill. 2010. On the shape and fabric of human history. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365(1559). 3923–3933. <https://doi.org/10.1098/rstb.2010.0162>.
- Gray, Russell D., Alexei J. Drummond & Simon J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323(5913). 479–483. <https://doi.org/10.1126/science.1166858>.
- Gray, Russell D., Simon J. Greenhill & Quentin D. Atkinson. 2013. Phylogenetic models of language change. In Peter J. Richerson, Morten H. Christiansen & Julia Lupp (eds.), *Cultural evolution: Society, technology, language, and religion*, 285–302. Cambridge: MIT Press. <https://doi.org/10.7551/mitpress/9780262019750.003.0015>.
- Greenhill, Simon J., Quentin D. Atkinson, Adam Meade & Russell D. Gray. 2010. The shape and tempo of language evolution. *Proceedings of the Royal Society B: Biological Sciences* 277(1693). 2443–2450. <https://doi.org/10.1098/rspb.2010.0051>.
- Greenhill, Simon J., Thomas E. Currie & Russell D. Gray. 2009. Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society B: Biological Sciences* 276(1665). 2299–2306. <https://doi.org/10.1098/rspb.2008.1944>.
- Greenhill, Simon J., Hannah J. Haynie, Robert M. Ross, Angela M. Chira, Johann-Mattis List, Lyle Campbell, Carlos A. Botero & Russell D. Gray. 2023. A recent northern origin for the Uto-Aztecan family. *Language*. <https://doi.org/10.1353/lan.0.0276>.
- Greenhill, Simon J., Paul Heggarty & Russell D. Gray. 2020. Bayesian phylolinguistics. In Richard D. Janda, Brian D. Joseph & Barbara S. Vance (eds.), *The handbook of historical linguistics*, vol. 2, 226–253. Hoboken: Wiley. <https://doi.org/10.1002/9781118732168.ch11>.
- Greenhill, Simon J., Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C. Levinson & Russell D. Gray. 2017. Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences* 114(42), E8822–E8829. <https://doi.org/10.1073/pnas.1700388114>.
- Grollemund, Rebecca, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti & Mark Pagel. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences* 112(43). 13296–13301. <https://doi.org/10.1073/pnas.1503793112>.
- Hagège, Claude & André Haudricourt. 1978. *La phonologie panchronique: Comment les sons changent dans les langues*. Paris: Presses Universitaires de France.
- Hartmann, Frederik. 2023. *Germanic phylogeny*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198872733.001.0001>.
- Haspelmath, Martin. 2004. How hopeless is genealogical linguistics, and how advanced is areal linguistics? *Studies in Language* 28(1). 209–223. <https://doi.org/10.1075/sl.28.1.10has>.

- Haynie, Hannah J. & Claire Bowern. 2016. Phylogenetic approach to the evolution of color term systems. *Proceedings of the National Academy of Sciences* 113(48). 13666–13671. <https://doi.org/10.1073/pnas.1613666113>.
- Heggarty, Paul, Cormac Anderson, Matthew Scarborough, Benedict King, Remco Bouckaert, Lechosław Jocz, Martin Joachim Kümmel, Thomas Jügel, Britta Irslinger, Roland Pooth, Henrik Liljegren, Richard F. Strand, Geoffrey Haig, Martin Macák, Ronald I. Kim, Erik Anonby, Tjimen Pronk, Oleg Belyaev, Tonya Kim Dewey-Findell, Matthew Boutilier, Cassandra Freiberg, Robert Tegethoff, Matilde Serangeli, Nikos Liosis, Krzysztof Stroński, Kim Schulte, Ganesh Kumar Gupta, Wolfgang Haak, Johannes Krause, Quentin D. Atkinson, Simon J. Greenhill, Denise Kühnert & Russell D. Gray. 2023. Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. *Science* 381(6656), eabg0818. <https://doi.org/10.1126/science.abg0818>.
- Hennig, Willi. 1950. *Grundzüge einer Theorie der phylogenetischen Systematik*. Berlin: Deutscher Zentralverlag.
- Hennig, Willi. 1979. *Phylogenetic systematics*. Urbana: University of Illinois Press.
- Hill, Jane H. 2011. Subgrouping in Uto-Aztecan. *Language Dynamics and Change* 1(2). 241–278. <https://doi.org/10.1163/221058212X643978>.
- Hoffmann, Konstantin, Remco Bouckaert, Simon J. Greenhill & Denise Kühnert. 2021. Bayesian phylogenetic analysis of linguistic data using BEAST. *Journal of Language Evolution* 6(2). 119–135. <https://doi.org/10.1093/jole/lzab005>.
- Hübschmann, Heinrich. 1897. *Armenische Grammatik*. Vol. 1: *Armenische Etymologie*. 2 vols. Leipzig: Breitkopf & Härtel.
- Jacques, Guillaume & Johann-Mattis List. 2019. Save the trees: Why we need tree models in linguistic reconstruction (and when we should apply them). *Journal of Historical Linguistics* 9(1). 128–166. <https://doi.org/10.1075/jhl.17008.mat>.
- Jacques, Guillaume & Thomas Pellard. 2021. Phylogenies based on lexical innovations refute the Rung hypothesis. *Diachronica* 38(1). 1–24. <https://doi.org/10.1075/dia.19058.jac>.
- Joseph, Brian D. 2012. Lexical diffusion and the regular transmission of language change in its sociohistorical context. In Juan Manuel Hernández-Campoy & Juan Camilo Conde-Silvestre (eds.), *The handbook of historical sociolinguistics*, 408–426. Chichester: Blackwell. <https://doi.org/10.1002/9781118257227.ch22>.
- Kaiping, Gereon A. & Marian Klamer. 2022. The dialect chain of the Timor-Alor-Pantar language family: A new analysis using systematic Bayesian phylogenetics. *Language Dynamics and Change* 12(2). 274–326. <https://doi.org/10.1163/22105832-bja10019>.
- Kalyan, Siva & Alexandre François. 2018. Freeing the comparative method from the tree model: A framework for Historical Glottometry. *Senri Ethnological Studies* 98. 59–89. <https://doi.org/10.15021/00009005>.
- Kalyan, Siva & Alexandre François. 2019. When the waves meet the trees: A response to Jacques and List. *Journal of Historical Linguistics* 9(1). 167–176. <https://doi.org/10.1075/jhl.18019.kal>.
- Kelly, Luke J. & Geoff K. Nicholls. 2017. Lateral transfer in stochastic Dollo models. *Annals of Applied Statistics* 11(2). 1146–1168. <https://doi.org/10.1214/17-A0AS1040>.
- Kluge, Friedrich. 1884. Die germanische Consonantendehnung. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 9(2). 149–186. <https://doi.org/10.1515/bgsl.1884.9.2.149>.
- Koile, Ezequiel, Simon J. Greenhill, Damián E. Blasi, Remco Bouckaert & Russell D. Gray. 2022. Phylogeographic analysis of the Bantu language expansion supports a rainforest route. *Proceedings*

- of the *National Academy of Sciences* 119(32), e2112853119. <https://doi.org/10.1073/pnas.2112853119>.
- Kolipakam, Vishnupriya, Fiona M. Jordan, Michael Dunn, Simon J. Greenhill, Remco Bouckaert, Russell D. Gray & Annemarie Verkerk. 2018. A Bayesian phylogenetic study of the Dravidian language family. *Royal Society Open Science* 5(3), 171504. <https://doi.org/10.1098/rsos.171504>.
- Kroonen, Guus. 2011. *The proto-Germanic n-stems: A study in diachronic morphophonology*. Amsterdam: Rodopi. <https://doi.org/10.1163/9789042032934>.
- Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic*. Leiden: Brill.
- Kümmel, Martin Joachim. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert.
- Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World lexicon of grammaticalization*. 2nd edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316479704>.
- Labov, William. 2007. Transmission and diffusion. *Language* 83(2). 344–387. <https://doi.org/10.1353/lan.2007.0082>.
- Lazaridis, Iosif, Songül Alpaslan-Roodenberg, Ayşe Acar, Ayşen Açikkol, Anagnostis Agelarakis, Levon Aghikyan, Uğur Akyüz, Desislava Andreeva, Gojko Andrijašević, Dragana Antonović, Ian Armit, Alper Atmaca, Pavel Avetisyan, Ahmet İhsan Aytek, Krum Bacvarov, Ruben Badalyan, Stefan Bakardzhiev, Jacqueline Balen, Lorenc Bejko, Rebecca Bernardos, Andreas Bertatos, Hanifi Biber, Ahmet Bilir, Mario Bodružić, Michelle Bonogofsky, Clive Bonsall, Dušan Borić, Nikola Borovinić, Guillermo Bravo Morante, Katharina Buttinger, Kim Callan, Francesca Candilio, Mario Carić, Olivia Cheronet, Stefan Chohadzhiev, Maria-Eleni Chovalopoulou, Stella Chrysoulaki, Ion Ciobanu, Natalija Čondić, Mihai Constantinescu, Emanuela Cristiani, Brendan J. Culleton, Elizabeth Curtis, Jack Davis, Ruben Davtyan, Tatiana I. Demcenco, Valentin Dergachev, Zafer Derin, Sylvia Deskaj, Seda Devejyan, Vojislav Djordjević, Kellie Sara Duffett Carlson, Laurie R. Eccles, Nedko Elenski, Atilla Engin, Nihat Erdoğan, Sabiha Erir-Pazarcı, Daniel M. Fernandes, Matthew Ferry, Suzanne Freilich, Alin Frînculeasa, Michael L. Galaty, Beatriz Gamarra, Boris Gasparyan, Bisserka Gaydarska, Elif Genç, Timur Gültekin, Serkan Gündüz, Tamás Hajdu, Volker Heyd, Suren Hobosyan, Nelli Hovhannisyán, Iliya Iliev, Lora Iliev, Stanislav Iliev, İlkyay İvgin, Ivor Janković, Lence Jovanova, Panagiotis Karkanias, Berna Kavaz-Kındıgılı, Esra Hilal Kaya, Denise Keating, Douglas J. Kennett, Seda Deniz Kesici, Anahit Khudaverdyan, Krisztián Kiss, Sinan Kılıç, Paul Klostermann, Sinem Kostak Boca Negra Valdes, Saša Kovačević, Marta Krenz-Niedbała, Maja Krznarić Škrivanko, Rovena Kurti, Pasko Kuzman, Ann Marie Lawson, Catalin Lazar, Krassimir Leshtakov, Thomas E. Levy, Ioannis Lirizis, Kirsi O. Lorentz, Sylwia Łukasik, Matthew Mah, Swapan Mallick, Kirsten Mandl, Kristine Martirosyan-Olshansky, Roger Matthews, Wendy Matthews, Kathleen McSweeney, Varduhi Melikyan, Adam Micco, Megan Michel, Lidija Milašinović, Alissa Mittnik, Janet M. Monge, Georgi Nekhrizov, Rebecca Nicholls, Alexey G. Nikitin, Vassil Nikolov, Mario Novak, Iñigo Olalde, Jonas Oppenheimer, Anna Osterholtz, Celal Özdemir, Kadir Toykan Özdoğan, Nurettin Öztürk, Nikos Papadimitriou, Niki Papakonstantinou, Anastasia Papathanasiou, Lujana Paraman, Evgeny G. Paskary, Nick Patterson, Ilian Petrakiev, Levon Petrosyan, Vanya Petrova, Anna Philippa-Touchais, Ashot Piliposyan, Nada Pocuca Kuzman, Hrvoje Potrebica, Bianca Preda-Bălănică, Zrinka Premužić, T. Douglas Price, Lijun Qiu, Siniša Radović, Kamal Raeuf Aziz, Petra Rajić Šikanjić, Kamal Rasheed Raheem, Sergei Razumov, Amy Richardson, Jacob Roodenberg, Rudenc Ruka, Victoria Russeva, Mustafa

- Şahin, Ayşegül Şarbak, Emre Savaş, Constanze Schattke, Lynne Schepartz, Tayfun Selçuk, Ayla Sevim-Erol, Michel Shamooun-Pour, Henry M. Shephard, Athanasios Sideris, Angela Simalcsik, Hakob Simonyan, Vitalij Sinika, Kendra Sirak, Ghenadie Sirbu, Mario Šlaus, Andrei Soficaru, Bilal Söğüt, Arkadiusz Sołtysiak, Çilem Sönmez-Sözer, Maria Stathi, Martin Steskal, Kristin Stewardson, Sharon Stocker, Fadime Suata-Alpaslan, Alexander Suvorov, Anna Szécsényi-Nagy, Tamás Szeniczey, Nikolai Telnov, Strahil Temov, Nadezhda Todorova, Ulsi Tota, Gilles Touchais, Sevi Triantaphyllou, Atila Türker, Marina Ugarković, Todor Valchev, Fanica Veljanovska, Zlatko Videvski, Cristian Virag, Anna Wagner, Sam Walsh, Piotr Włodarczyk, J. Noah Workman, Aram Yardumian, Evgenii Yarovoy, Alper Yener Yavuz, Hakan Yılmaz, Fatma Zalzala, Anna Zettl, Zhao Zhang, Rafet Çavuşoğlu, Nadin Rohland, Ron Pinhasi & David Reich. 2022. The genetic history of the Southern Arc: A bridge between West Asia and Europe. *Science* 377(6609), eabm4247. <https://doi.org/10.1126/science.abm4247>.
- Lehtinen, Jyri, Terhi Honkola, Kalle Korhonen, Kaj Syrjänen, Niklas Wahlberg & Outi Vesakoski. 2014. Behind family trees: Secondary connections in Uralic language networks. *Language Dynamics and Change* 4(2). 189–221. <https://doi.org/10.1163/22105832-00402007>.
- Lemey, Philippe, Andrew Rambaut, John J. Welch & Marc A. Suchard. 2010. Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution* 27(8). 1877–1885. <https://doi.org/10.1093/molbev/msq067>.
- Lemey, Philippe, Marco Salemi & Anne-Mieke Vandamme (eds.). 2009. *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing*. 2nd edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09780511819049>.
- Leskien, August. 1876. *Die Declination im Slavisch-Litauischen und Germanischen*. Leipzig: Hirzel. [http://www.deutschestextarchiv.de/book/show/leskien\\_declination\\_1876](http://www.deutschestextarchiv.de/book/show/leskien_declination_1876).
- List, Johann-Mattis. 2019. Automatic inference of sound correspondence patterns across multiple languages. *Computational Linguistics* 45(1). 137–161. [https://doi.org/10.1162/coli\\_a\\_00344](https://doi.org/10.1162/coli_a_00344).
- List, Johann-Mattis. 2022. Correcting a bias in TIGER rates resulting from high amounts of invariant and singleton cognate sets. *Journal of Language Evolution* 7(1). 53–58. <https://doi.org/10.1093/jole/lzab007>.
- List, Johann-Mattis, Simon J. Greenhill & Russell D. Gray. 2017. The potential of automatic word comparison for historical linguistics. *PLOS One* 12(1), e0170046. <https://doi.org/10.1371/journal.pone.0170046>.
- List, Johann-Mattis, Shijulal Nelson-Sathi, Hans Geisler & William Martin. 2014. Networks of lexical borrowing and lateral gene transfer in language and genome evolution. *BioEssays* 36(2). 141–150. <https://doi.org/10.1002/bies.201300096>.
- List, Johann-Mattis, Jananan Sylvestre Pathmanathan, Philippe Lopez & Eric Baptiste. 2016. Unity and disunity in evolutionary sciences: Process-based analogies open common research avenues for biology and linguistics. *Biology Direct* 11(1). 39. <https://doi.org/10.1186/s13062-016-0145-2>.
- Macklin-Cordes, Jayden L., Claire Bowerman & Erich R. Round. 2021. Phylogenetic signal in phonotactics. *Diachronica* 38(2). 210–258. <https://doi.org/10.1075/dia.20004.mac>.
- Macklin-Cordes, Jayden L. & Erich R. Round. 2022. Challenges of sampling and how phylogenetic comparative methods help: With a case study of the Pama-Nyungan laminal contrast. *Linguistic Typology* 26(3). 533–572. <https://doi.org/10.1515/lingty-2021-0025>.
- Maddison, Wayne P. 1997. Gene trees in species trees. *Systematic Biology* 46(3). 523–536. <https://doi.org/10.1093/sysbio/46.3.523>.

- Matsumae, Hiromi, Peter Ranacher, Patrick E. Savage, Damián E. Blasi, Thomas E. Currie, Kae Koganebuchi, Nao Nishida, Takehiro Sato, Hideyuki Tanabe, Atsushi Tajima, Steven Brown, Mark Stoneking, Kentaro K. Shimizu, Hiroki Oota & Balthasar Bickel. 2021. Exploring correlations in genetic and cultural variation across language families in northeast Asia. *Science Advances* 7(34), eabd9223. <https://doi.org/10.1126/sciadv.abd9223>.
- McMahon, April & Robert McMahon. 2006. Why linguists don't do dates: Evidence from Indo-European and Australian languages. In Peter Forster & Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, 153–160. Cambridge: McDonald Institute for Archaeological Research.
- Meillet, Antoine. 1925. *La méthode comparative en linguistique historique*. Oslo, Leipzig & Paris: H. Aschehoug, Otto Harrassowitz & Honoré Champion. <http://gallica.bnf.fr/ark:/12148/bpt6k934150p>.
- Morrison, David. 2011. *Introduction to phylogenetic networks*. Uppsala: RJR Productions.
- Münkemüller, Tamara, Sébastien Lavergne, Bruno Bzeznik, Stéphane Dray, Thibaut Jombart, Katja Schiffers & Wilfried Thuiller. 2012. How to measure and test phylogenetic signal. *Methods in Ecology and Evolution* 3(4). 743–756. <https://doi.org/10.1111/j.2041-210X.2012.00196.x>.
- Neureiter, Nico, Peter Ranacher, Nour Efrat-Kowalsky, Gereon A. Kaiping, Robert Weibel, Paul Widmer & Remco R. Bouckaert. 2022. Detecting contact in language trees: A Bayesian phylogenetic model with horizontal transfer. *Humanities and Social Sciences Communications* 9(205). 1–14. <https://doi.org/10.1057/s41599-022-01211-7>.
- Neureiter, Nico, Peter Ranacher, Rik van Gijn, Balthasar Bickel & Robert Weibel. 2021. Can Bayesian phylogeography reconstruct migrations and expansions in linguistic evolution? *Royal Society Open Science* 8(1). <https://doi.org/10.1098/rsos.201079>.
- O'Reilly, Joseph E. & Philip C. J. Donoghue. 2018. The efficacy of consensus tree methods for summarizing phylogenetic relationships from a posterior sample of trees estimated from morphological data. *Systematic Biology* 67(2). 354–362. <https://doi.org/10.1093/sysbio/syx086>.
- O'Reilly, Joseph E., Mark N. Puttick, Luke Parry, Alastair R. Tanner, James E. Tarver, James Fleming, Davide Pisani & Philip C. J. Donoghue. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology Letters* 12(4), 20160081. <https://doi.org/10.1098/rsbl.2016.0081>.
- Pagel, Mark. 1999. Inferring the historical patterns of biological evolution. *Nature* 401(6756). 877–884. <https://doi.org/10.1038/44766>.
- Pagel, Mark, Quentin D. Atkinson & Andrew Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449(7163). 717–720. <https://doi.org/10.1038/nature06176>.
- Pagel, Mark, Andrew Meade & Daniel Barker. 2004. Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* 53(5). 673–684. <https://doi.org/10.1080/10635150490522232>.
- Pellard, Thomas. 2015. The linguistic archeology of the Ryukyu Islands. In Patrick Heinrich, Shinsho Miyara & Michinori Shimoji (eds.), *Handbook of the Ryukyuan languages: History, structure, and use*, 13–37. Berlin: De Gruyter. <https://doi.org/10.1515/9781614511151.13>.
- Phillips, Joshua & Claire Bowern. 2022. Bayesian methods for ancestral state reconstruction in morphosyntax: Exploring the history of argument marking strategies in a large language family. *Journal of Language Evolution* 7(1). 1–15. <https://doi.org/10.1093/jole/lzac002>.
- Potochnik, Angela. 2017. *Idealization and the aims of science*. Chicago: University of Chicago Press.

- Puttick, Mark N., Joseph E. O'Reilly, Alastair R. Tanner, James F. Fleming, James Clark, Lucy Holloway, Jesus Lozano-Fernandez, Luke A. Parry, James E. Tarver, Davide Pisani & Philip C. J. Donoghue. 2017. Uncertain-tree: Discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of the Royal Society B: Biological Sciences* 284(1846), 20162290. <https://doi.org/10.1098/rspb.2016.2290>.
- Rix, Helmut, Martin Kümmel, Thomas Zehnder, Reiner Lipp & Brigitte Schirmer. 2001. *Lexikon der indogermanischen Verben: Die Wurzeln und ihre Primärstambildungen*. 2nd edn. Wiesbaden: Reichert.
- Robert, Christian P. 2007. *The Bayesian choice: From decision-theoretic foundations to computational implementation*. 2nd edn. New York: Springer. <https://doi.org/10.1007/0-387-71599-1>.
- Robins, R. H. 1973. The history of language classification. In Thomas A. Sebeok, Henry M. Hoenigswald & Robert E. Longacre (eds.), *Diachronic, areal, and typological linguistics*, 3–42. The Hague: Mouton. <https://doi.org/10.1515/9783111418797-003>.
- Ronquist, Fredrik, Paul van der Mark & John P. Huelsenbeck. 2009. Bayesian phylogenetic analysis using MrBayes. In Philippe Lemey, Marco Salemi & Anne-Mieke Vandamme (eds.), *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing*, 2nd edn., 210–266. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09780511819049.009>.
- Ross, Malcom D. 1997. Social networks and kinds of speech-community event. In Roger Blench & Matthew Spriggs (eds.), *Archaeology and language I: Theoretical and methodological orientations*, 209–261. London: Routledge.
- Round, Erich, Rikker Dockum & Robin J. Ryder. 2022. Evolution and trade-off dynamics of functional load. *Entropy* 24(4), 507. <https://doi.org/10.3390/e24040507>.
- Ryder, Robin. 2010. *Phylogenetic models of language diversification*. Oxford: University of Oxford dissertation. <https://theses.hal.science/tel-00661866>.
- Ryder, Robin J. & Geoff K. Nicholls. 2011. Missing data in a stochastic Dollo model for binary trait data, and its application to the dating of Proto-Indo-European. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 60(1). 71–92. <https://doi.org/10.1111/j.1467-9876.2010.00743.x>.
- Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2). 411–444. <https://doi.org/10.1353/ol.2005.0012>.
- Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin J. Ryder, Valentin Thouzeau, Simon J. Greenhill & Johann-Mattis List. 2019. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences* 116(21). 10317–10322. <https://doi.org/10.1073/pnas.1817972116>.
- Saitou, Naruya & Masatoshi Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4(4). 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>.
- Sangsrgyas Tshering. 2020. Diébù Zàngyǔ de qīngyīn zhuóhuà. *Cahiers de linguistique Asie orientale* 49(1). 1–20. <https://doi.org/10.1163/19606028-bja10001>.
- Sankoff, David. 1973. Mathematical developments in lexicostatistic theory. In Thomas A. Sebeok, Henry M. Hoenigswald & Robert E. Longacre (eds.), *Diachronic, areal, and typological linguistics*, 93–113. The Hague: Mouton. <https://doi.org/10.1515/9783111418797>.
- Sankoff, David & Gillian Sankoff. 1980. Wave versus *Stammbaum* explanations of lexical similarities. In *The social life of language*, 143–152. Philadelphia: University of Pennsylvania Press. <https://doi.org/10.9783/9781512809589-010>.

- Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace and Company.
- Schleicher, August. 1861–1862. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. 2 vols. Weimar: Hermann Böhlau.
- Schmidt, Heiko A. & Arndt von Haeseler. 2009. Phylogenetic inference using maximum likelihood methods. In Philippe Lemey, Marco Salemi & Anne-Mieke Vandamme (eds.), *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing*, 2nd edn., 181–209. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09780511819049.008>.
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Hermann Böhlau. <http://dl.ub.uni-freiburg.de/diglit/schmidt1872>.
- Schuchardt, Hugo. 1900. *Über die Klassifikation der romanischen Mundarten: Probevorlesung gehalten zu Leipzig am 30. april 1870*. Graz: Styria. <https://archive.org/embed/berdieklassifik00schugoog>.
- Sebeok, Thomas A., Henry M. Hoenigswald & Robert E. Longacre (eds.). 1973. *Diachronic, areal, and typological linguistics*. The Hague: Mouton. <https://doi.org/10.1515/9783111418797>.
- Seebold, Elmar. 1970. *Vergleichendes und etymologisches Wörterbuch der germanischen starken Verben*. The Hague: Mouton. <https://doi.org/10.1515/9783110821956>.
- Sokal, Robert R. & Charles Duncan Michener. 1958. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 38(22). 1409–1438. <https://www.biodiversitylibrary.org/page/3711319>.
- Swadesh, Morris. 1952. Lexico-statistic dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96(4). 452–463. <https://www.jstor.org/stable/3143802>.
- Swofford, David L. & Jack Sullivan. 2009. Phylogeny inference based on parsimony and other methods using PAUP\*. In Philippe Lemey, Marco Salemi & Anne-Mieke Vandamme (eds.), *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing*, 2nd edn., 267–312. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09780511819049.010>.
- Syrjänen, Kaj, Luke Maurits, Unni Leino, Terhi Honkola, Jadranka Rota & Outi Vesakoski. 2021. Crouching TIGER, hidden structure: Exploring the nature of linguistic data using TIGER values. *Journal of Language Evolution* 6(2). 99–118. <https://doi.org/10.1093/jole/lzab004>.
- Tian, Zheng, Yuxin Tao, Kongyang Zhu, Guillaume Jacques, Robin J. Ryder, José Andrés Alonso de la Fuente, Anton Antonov, Ziyang Xia, Yuxuan Zhang, Xiaoyan Ji, Xiaoying Ren, Guanglin He, Jianxin Guo, Rui Wang, Xiaomin Yang, Jing Zhao, Dan Xu, Russell D. Gray, Menghan Zhang, Shaoqing Wen, Chuan-Chao Wang & Thomas Pellard. 2022. Triangulation fails when neither linguistic, genetic, nor archaeological data support the Transeurasian narrative. *bioRxiv*. <https://doi.org/10.1101/2022.06.09.495471>.
- Traugott, Elizabeth Closs & Richard B. Dasher. 2001. *Regularity in semantic change*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09780511486500>.
- Trubetzkoy, Nikolai S. 1939. Gedanken über das Indogermanenproblem. *Acta Linguistica* 1(1). 81–89. <https://doi.org/10.1080/03740463.1939.10410851>.
- Van de Peer, Yves & Marco Salemi. 2009. Phylogenetic inference based on distance methods. In Philippe Lemey, Marco Salemi & Anne-Mieke Vandamme (eds.), *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing*, 2nd edn., 142–180. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09780511819049.007>.
- Velasco, Joel D. 2012. The future of systematics: Tree thinking without the tree. *Philosophy of Science* 79(5). 624–636. <https://doi.org/10.1086/667878>.

- Verkerk, Annemarie. 2017. Phylogenies: Future, not fallacy. *Language Dynamics and Change* 7(1). 127–140. <https://doi.org/10.1163/22105832-00601013>.
- Wichmann, Søren & Taraka Rama. 2021. Testing methods of linguistic homeland detection using synthetic data. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376(1824), 20200202. <https://doi.org/10.1098/rstb.2020.0202>.
- Winter, Werner. 1973. Areal linguistics: Some general considerations. In Thomas A. Sebeok, Henry M. Hoenigswald & Robert E. Longacre (eds.), *Diachronic, areal, and typological linguistics*, 135–148. The Hague: Mouton. <https://doi.org/10.1515/9783111418797-009>.
- Wodtko, Dagmar S., Britta Irslinger & Carolin Schneider. 2008. *Nomina im indogermanischen Lexikon*. Heidelberg: Universitätsverlag Winter.
- Zhang, Menghan, Shi Yan, Wuyun Pan & Li Jin. 2019. Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature* 569. 112–115. <https://doi.org/10.1038/s41586-019-1153-z>.
- Zhou, Kevin & Claire Bowern. 2015. Quantifying uncertainty in the phylogenetics of Australian numeral systems. *Proceedings of the Royal Society B: Biological Sciences* 282(1815), 20151278. <https://doi.org/10.1098/rspb.2015.1278>.