

Title:

Detecting non-tree-like signal using multiple tree topologies

Author:

Annemarie Verkerk

This is the accepted manuscript of an article published in *Journal of Historical Linguistics*, Volume 9 (2019), Issue 1, p. 9-69. The version of record is available online at <https://doi.org/10.1075/jhl.17009.ver>

The article is protected by copyright. The publisher John Benjamins Publishing Company should be contacted for permission to re-use or reprint the material in any form.

Citation:

Verkerk, Annemarie (2019): Detecting non-tree-like signal using multiple tree topologies. *Journal of Historical Linguistics*, 9(1), 9-69. DOI: 10.1075/jhl.17009.ver



Detecting non-tree-like signal using multiple tree topologies

Annemarie Verkerk
University of Reading &
Max Planck Institute for the Science of Human History

Abstract: Recent applications of phylogenetic methods to historical linguistics have been criticized for assuming a tree structure in which ancestral languages differentiate and split up into daughter languages, while language evolution is inherently non-tree-like (François 2014; Blench 2015: 32–33). This article attempts to contribute to this debate by discussing the use of the multiple topologies method (Pagel & Meade 2006a) implemented in *BayesPhylogenies* (Pagel & Meade 2004). This method is applied to lexical datasets from four different language families: Austronesian (Gray, Drummond & Greenhill 2009), Sinitic (Ben Hamed & Wang 2006), Indo-European (Bouckaert et al. 2012), and Japonic (Lee & Hasegawa 2011). Evidence for multiple topologies is found in all families except, surprisingly, Austronesian. It is suggested that reticulation may arise from a number of processes, including dialect chain break-up, borrowing (both shortly after language splits and later on), incomplete lineage sorting, and characteristics of lexical datasets. It is shown that the multiple topologies method is a useful tool to study the dynamics of language evolution.

Keywords: Bayesian phylogenetic inference, Austronesian, Sinitic, Indo-European, Japonic, language contact, reticulation

1. Introduction

In the last eighteen years, phylogenetic methods from evolutionary biology have made inroads into historical linguistics: these methods are applied both to building phylogenetic trees (Ben Hamed & Wang 2006; Gray, Drummond & Greenhill 2009; Lee & Hasegawa 2011; Bouckaert et al. 2012; Grollemund et al. 2015) and to inferring the process of evolution of typological characteristics on trees (Dunn et al. 2011; Verkerk 2014; Zhou & Bowerman 2015). The reception of these studies has been mixed, with studies criticizing data type (generally cognate-coded lexical data), data quality, the applicability of methods and models from another discipline to linguistic data, and the limited possibilities for incorporating previous knowledge into the phylogenetic analysis (Eskaya & Ringe 2004; Heggarty 2006; Holm 2007; Blench 2015; Pereltsvaig & Lewis 2015).

Nevertheless, phylogenetic methods have been adopted by a wide range of linguists to answer questions regarding the diversification of language families and of typological features (Bowerman & Atkinson 2012; Bouchard-Côté et al. 2013; Galucio et al. 2015; Macklin-Cordes & Round 2015; Meira, Birchall & Chousou-Polydouri 2015; Jaeger & Wichmann 2016; Widmer et al. 2017). Dunn (2014: 191), in his chapter in Bowerman & Epps' (2014) *The Routledge Handbook of Historical Linguistics*, notes that they are “increasingly well received within linguistics.” This positive reception is for good reason, given that evolutionary biologists have been considering

statistical approaches to the study of species diversification for over fifty years now: the first maximum likelihood analysis of (human) genetic material came from Edwards & Cavalli-Sforza (1964). Dunn (2014) describes how linguists began considering quantitative approaches to language history only when Swadesh (Swadesh 1952, 1955) first developed his methods of lexicostatistics and glottochronology, with recent growth having occurred after the adoption of methods from evolutionary biology by Gray & Jordan (2000). Thus, the methods of evolutionary biology are likely to complement those of traditional historical linguistics, especially when it comes to quantitative methods for inferring language genealogies and changes in cultural and linguistic features in these genealogies (Pagel 2009; Levinson & Gray 2012).

In the current article, a phylogenetic inference technique taken from biology is applied to linguistic data, and its usability is reviewed. The technique is called “multiple tree topologies.” It is implemented using the software *BayesPhylogenies*, which provides a Bayesian framework for inferring trees for a variety of data types, including binary cognate-coded lexical data (Pagel & Meade 2004). The multiple tree topologies method does not assume that the data have a single evolutionary history, but instead fits multiple independent tree topologies to the data. If, for whatever reason, certain sites (columns in the dataset, which here constitute the cognate sets) point towards a different cladistic grouping than others do, these different signals are picked up and then reflected by displaying statistical support for two or more different tree topologies which are estimated simultaneously. This method is applied to lexical datasets from four language families: Austronesian (Gray, Drummond & Greenhill 2009), Sinitic (Ben Hamed & Wang 2006), Indo-European (Bouckaert et al. 2012), and Japonic (Lee & Hasegawa 2011). This article contributes to the use of phylogenetic methods in historical linguistics by applying a specific phylogenetic method to linguistic data. It is especially relevant to the question of whether phylogenetic tree inference, which models change in contemporary entities in terms of descent with modification from common ancestors, can deal with non-tree-like aspects of language change.

To perform a strict test of this method, we must apply it to a dataset that we know includes both (i) genealogically inherited characteristics that can be traced to a process of descent with modification and (ii) features that have arisen through other processes, including dialect chain break-up, borrowing (from a substrate, superstrate, or adstrate), or incomplete lineage sorting. After applying the multiple topologies method, we can then assess how much of the tree-like and non-tree-like signal is recovered and correctly characterized. However, this required dataset does not exist, as the bundles of data that are studied by historical linguistics contain features that could have arisen via any of these processes (Heggarty, Maguire & McMahon 2010: 3829), and quantitative methods for exploring a dataset in terms of these processes are in their infancy or non-existent (Huson & Scornavacca 2010: 29). In some cases, it is possible to distinguish different sources of divergent signals on the basis of knowledge about which features are more likely to arise through a particular process or through detailed knowledge of surrounding languages or families, but these are not strict tests. A strict test of the multiple topologies method or of any method that attempts to incorporate multiple historical signals in some way is therefore not possible at this point in time.

In this light, this study should be seen as an exploratory experiment for the purpose of observing and reporting on the behavior of the multiple topologies method. However, as this article is part of this special issue exploring non-cladistic approaches to language genealogy, the study is also embedded in a larger context of investigating reticulation patterns in language and culture. Within the context of this discussion, after presenting the results from each of the four language families, I speculate on the processes that might have given rise to the evidence found for multiple topologies in three of the language families. My speculation is based on the assumption that if evidence for multiple topologies is found, and if a language X is found in different

language groupings in each tree, this implies that the lexical dataset of language X contains evidence for its affiliation with different languages through different historical processes. If the only relevant process of language change was genealogical descent with modification, such mixed signals would not be found. (This does not mean that absence of evidence for multiple topologies means that genealogical descent with modification is the only relevant process of language change.) Therefore, I draw on horizontal transmission processes as well as incomplete lineage sorting to explain evidence for multiple topologies. Whether this assumption is valid or not is a matter of further theorizing and strict tests that are beyond the limits of the current article.

2. Reticulation in language and culture

Despite the growing acceptance of phylogenetic methods in historical linguistics, it is important to keep in mind their limitations in order to improve upon them if possible. One criticism of phylogenetic methods with both practical and theoretical significance is that language change is not tree-like (Blench 2015; François 2014; Geisler & List 2013; see also references in Atkinson & Gray 2005:523 and Croft 2008:228–229). Languages, like any part of culture, do not evolve only through an ancestral language differentiating and splitting up into daughter languages which then diverge and split further; rather, they are also characterized by features taken from languages other than their parents. This is problematic for phylogenetic tree inference if rates of borrowing are high – imagine if language A borrowed features from a related language B and these features are subsequently inherited by the descendants of language A. The descendants of language A will appear more similar to language B and its descendants because of the borrowed features (see Figure 1 for some terminology). This may have an impact on the placement of languages A and B, along with their descendants, in a phylogenetic tree, causing them to be grouped more closely together (Wang & Minett 2005). A case in point is the position of English in the Indo-European family tree, as exemplified in Figure 1. English has borrowed extensively from Latin, North Germanic, and (Norman) French (Algeo & Pyles 2005:247–268). Taking into account none of the borrowings but only its history within Germanic, the true genealogical position of English is as a sister language to Frisian (Beekes 2011: 29); however, the phylogenetic analysis of Indo-European languages by Bouckaert et al. (2012), displayed with some changes in Figure 1, places it outside of the West Germanic branch.

The non-tree-like character of language change is reflected not only by borrowing after languages have split, but also by the processes through which new languages come into being. François (2014) and Kalyan & François (2018) make a case against the tree model from the perspective of the wave model of language change, which focuses on dialect networks. They argue that the tree model fails to capture what is perhaps the most common language diversification process, namely the “fragmentation of a language into a network of dialects which remained in contact with each other for an extended period of time” (François 2014: 163). This scenario is not well represented by a tree structure, as trees cannot deal with the reticulation introduced by extended (and perhaps continued) language contact and borrowing by emerging language clades. Typically, isoglosses do not form neat linguistic subgroups, but show widely varying overlaps.

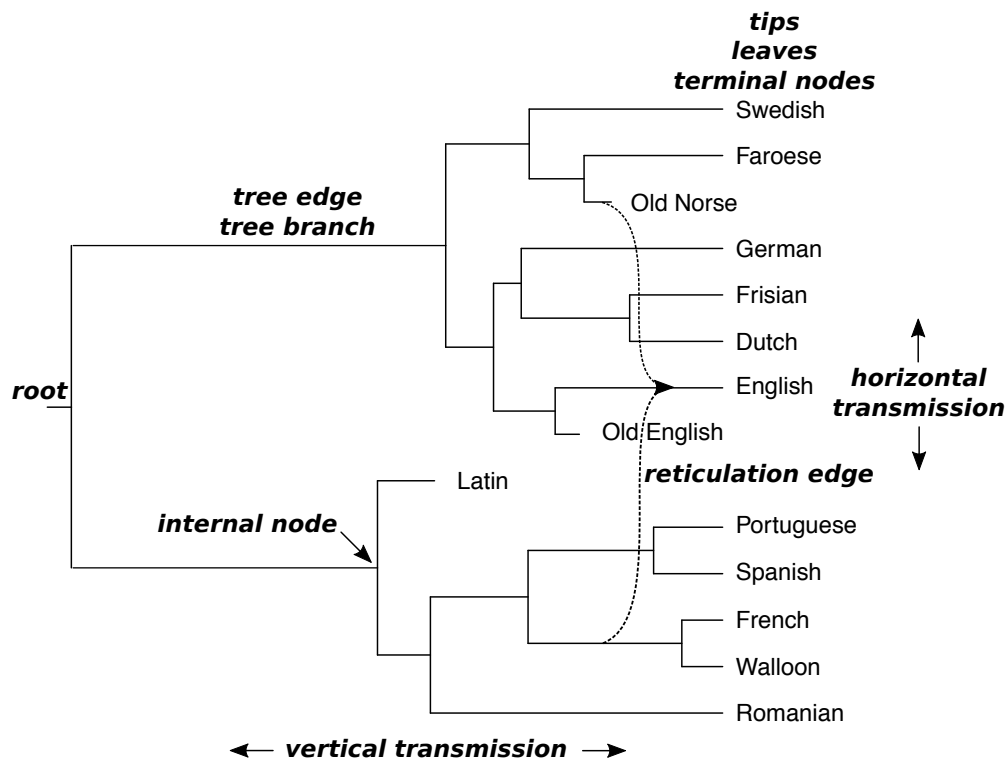


Figure 1. A simplified rooted phylogenetic tree of some Germanic and some Romance languages, illustrating borrowings from the ancestors of the modern Scandinavian languages and French into English during the Old and Middle English period¹

This reflects the situation in evolutionary biology, as well, where divergences between gene histories and species histories are well known (see Gray, Greenhill & Ross 2007: 367–368 and List et al. 2016 for these processes in linguistics). Maddison (1997: 523) describes a phylogenetic tree of species as “a cloud of gene histories.” Gene trees may not match with species trees because of horizontal transfer (hybridization), incomplete lineage sorting, and gene duplication and extinction (Maddison 1997; see Koonin, Makarova & Aravind 2001 for an introduction to horizontal gene transfer). Incomplete lineage sorting occurs when there are ancestral polymorphisms, which are by chance inherited only partially by different species, suggesting a different history from the species tree (see List et al. 2016: 16 and Jacques & List, this issue). The problem with gene duplication and extinction is similar: if a gene duplicates in an ancestral species, the two duplicates embark on different evolutionary paths. If some of these copies subsequently go extinct, the true history of the duplicated gene cannot be recovered. Rokas & Carroll (2006: 1903) add long branch attraction to the list of processes affecting the recovery of the species tree: when branches for certain species are very long, subsequent changes will overwhelm the historical signal needed to retrieve the species tree (see Holm 2007: 185 for the same point in linguistics).

Processes that generate reticulation affect both biological and linguistic change. Hence, phylogenetic inference can be problematic not simply because words and other linguistic features can be borrowed across lineages, but also because in a wider sense, words have their own unique etymologies (like “gene trees” in biology) that may or may not match with the family tree (“species tree”). Additionally, words can change their denotatum easily, i.e., can become irrelevant for the study of a particular concept due to semantic shift. For example, the English word *hound* is irrelevant when we study Germanic words for the concept ‘dog’, while its relatives

¹ The tree topology is based on Bouckaert et al. (2012), a Bayesian phylogeographic analysis of cognate-coded lexical data. The labels indicate terms frequently used in linguistic phylogenetic studies.

in German (*Hund*) and Dutch (*hond*) are still relevant. Genes, on the other hand, cannot really change their function, as mutations in a given gene are likely to render the gene defunct or make it more efficient but are not likely to change its function altogether. Hence, we expect that linguistics is more greatly impacted by discrepancies between “gene trees” and “species trees.” On the other hand, note that if these problems were severe, phylogenies could not be inferred with any accuracy. This is not true: it is known from several well-studied language families such as Indo-European and Austronesian that it is possible to approximate tree topologies established by conventional techniques (Gray, Drummond & Greenhill 2009; Bouckaert et al. 2012, among others).

Researchers in linguistics, anthropology, and biology have considered several solutions to deal with reticulation, both within and outside of the tree model (please note that no attempt has been made to provide an exhaustive review of all these proposed solutions, as the literature on this topic is extensive). Within the tree model, several studies have simulated data with known rates of borrowing and subsequently tested the behavior of phylogenetic tree inference (Nunn, Borgerhoff Mulder & Langlely 2006; Greenhill, Currie & Gray 2009; Nunn et al. 2010). Collard, Shennan & Tehrani (2006) compared biological and cultural datasets using the retention index (RI), a goodness-of-fit measure, and showed that cultural change is neither more nor less tree-like than biological change. Matthews et al. (2011) tested two different models of evolution of Iranian tribal textiles on trees and found that various sets of techniques and designs have different histories of descent. Bowerman et al. (2011) showed that hunter-gatherer language lexicons, unlike previously thought, do not have higher rates of borrowing than agriculturalist languages. Tehrani & Collard (2002) demonstrated the importance of tree-like change over borrowing in Turkmen textiles.

Moving beyond the tree model, the most obvious way to incorporate and account for reticulation is through using phylogenetic networks. The case for using them has been made both in biology (Huson & Scornavacca 2010, 2012; Morrison 2016) and linguistics (List et al. 2013). Nakhleh, Ringe & Warnow (2005) presented a method for inferring “perfect phylogenetic networks” by adding a minimal number of lateral edges for characters that do not fit on a set of highly regarded trees. Heggarty, Maguire & McMahon (2010) discussed a network approach based on phonetic divergence. Wichmann et al. (2011) contributed a discussion of measures of reticulation (delta scores and Q-residuals). Nelson-Sathi et al. (2011) analyzed two Indo-European datasets using minimal lateral networks, which add lateral edges to a reference tree for cognate sets that do not fit the tree structure; they found that borrowing is pervasive. List et al. (2013:147) stated that Nelson-Sathi et al.’s (2011) results are overblown; they applied the same method and found that 369 out of 1,190 cognate sets (31%) were affected by borrowing, considerably less than the 61% found by Nelson-Sathi et al. (2011). Kelly & Nicholls (2017) proposed a stochastic Dollo model that includes lateral transfer in a Bayesian context and tested this model using Eastern Polynesian languages. The most recent application is Willems et al. (2016), who used hybridization networks with an Indo-European dataset; this approach had the advantage of being able to find the directionality of lateral transfer in terms of donor and recipient language. Many of these papers have provided overviews of even more applications of rooted network techniques in linguistics, suggesting that this is a growing area of inquiry.

This article is an exploratory study of a method that may be able to capture reticulation in linguistics. The multiple topologies method infers multiple tree topologies, rather than just one, and is informative with regard to the sites (columns in the dataset; in the current study, cognate sets) that support each topology. In this article, the set of trees inferred always equals two, although more are possible. This technique has the benefit of not requiring a reference tree of the sort used by Nakhleh, Ringe & Warnow (2005), which is required for minimal lateral

networks (Nelson-Sathi et al. 2011, List et al. 2013). It was developed in order to account for divergences between gene histories and species histories in evolutionary biology, and therefore it may be suited to account for similar patterns in linguistics. I describe the outcome of applying this tool on four lexical linguistic datasets in Section 4 and speculate on possible non-tree-like processes that could be responsible for the multiple topologies found in Section 5. The four lexical datasets are the following:

- i. Austronesian: Gray, Drummond & Greenhill (2009): 400 languages; 210 concepts; 34440 cognate sets. The Austronesian dataset is the largest one considered here, with 400 languages. The authors removed loans from the lexical dataset. Their best-performing model was a covarion model. They produced a dated tree that shows how the expansion of the Austronesian language family is coupled with the expansion of its speakers across the Pacific and that this expansion is probably linked to the development of sea-faring technologies.
- ii. Sinitic: Ben Hamed & Wang (2006): 24 languages; 200 concepts. This dataset was downloaded from the Chinese Dialect Database (List 2017), a repository for linguistic information on Chinese dialects, including several lexical datasets. In the database, the dataset used by Ben Hamed & Wang (2006) has been made available as “Wang (2004a),” cited in this article as Wang (2004). In this article I refer to their paper as “Ben Hamed & Wang (2006)” when I compare the results of this study to their phylogenetic results; Wang (2004) is only a reference to the dataset. The only difference between the dataset as used by Ben Hamed & Wang (2006) and the dataset available as “Wang (2004a)” from List (2017) is that the latter does not include Old Chinese. This language was excluded by List, as it should not be part of a contemporary dataset for use in studies of non-tree-like signal. For this reason, the dataset used in the current article contains 23 languages, 200 concepts, and 1511 cognate sets. Ben Hamed & Wang (2006) make clear that a fair amount of reticulation is present in the Sinitic dataset. Their article features an extensive discussion of the applicability of phylogenetic tree inference and network inference methods. Known borrowings were removed from the dataset. The same dataset was subsequently analyzed by List (2015, 2016). Further analysis on the Sinitic languages using a different dataset is performed by List et al. (2014).
- iii. Indo-European: Bouckaert et al. (2012): 103 languages; 207 concepts; 5997 cognate sets. The aim of Bouckaert et al. (2012) was to demonstrate an Anatolian origin for the Indo-European language family. In their dataset, known borrowings were removed. Their best performing model was the stochastic Dollo model; their preferred clock was the uncorrelated lognormal relaxed clock. Note that since their model included ancestral inferences of geography, this may have impacted which model and clock was best supported. Since it was not possible to link the columns of this binary dataset to the concepts of the cognate sets, Michael Dunn was kind enough to provide a dataset for which this was possible (Dunn 2018). This dataset can be considered a slightly revised version in terms of cognate coding. It contains 116 languages, 207 concepts, and 6381 cognate sets. The languages that are additional compared to the dataset used by Bouckaert et al. (2012) are listed in the relevant section below.
- iv. Japonic: Lee & Hasegawa (2011): 59 languages; 210 concepts; 675 cognate sets. The authors found that the covarion model in combination with a relaxed clock produced the best fit. They produce a dated tree. The main claim of their article is that the arrival of the Japonic languages coincided with the arrival of agriculture in what is now Japan. There is considerable uncertainty in their tree sample, however. While the higher internal nodes are well resolved, many lower internal nodes on the maximum clade credibility tree have a posterior probability of less than .5 (Lee & Hasegawa 2011:4). See Lee (2018) for the dataset and resulting tree set.

These four datasets were chosen because of their diverging features in terms of size and evolutionary history. Austronesian and Indo-European are large language families, and Indo-European especially is well-studied. Both of these families are of continued interest among phylogeneticists, and new phylogenetic analyses featuring more languages will be presented in the future. The Sinitic and Japonic datasets feature dialects and languages, but for the sake of convenience all language varieties that are differentiated are referred to as “languages.” These languages may involve more reticulation (a fact which is definitely true for the Sinitic dataset, as described by Ben Hamed & Wang 2006) and a different type of evolutionary history in comparison to Austronesian and Indo-European.

In these four datasets, known borrowings were removed (Lee & Hasegawa 2011:3667 do not explicitly state this, but suggest that they have done so in their discussion, in which they point out that knowing whether one has indeed removed all loans is impossible). The goal of this article is to explicitly investigate these datasets, because as both these papers and the results in Section 4 show, evidence for reticulation is still present despite removal of known loans. We know that languages are full of loanwords in general (Tadmor 2009); Nelson-Sathi et al. (2011) and List et al. (2013) have shown undetected borrowings to be present in earlier Indo-European lexical datasets, and I side with Lee & Hasegawa (2011) in agreeing that detecting borrowings can be very difficult (see List et al. 2016: 7 for perspectives on automatic loanword identification). Aside from borrowing, there are other processes, such as incomplete lineage sorting, that result in non-tree-like signal. Therefore, the current article is an application of the multiple topologies method to datasets typically used for phylogenetic inference, which as far as we know typically contain non-tree-like signal despite removal of known loanwords. It would additionally be useful to compare the current results to an analysis of the same dataset with known loans kept in, in order to see whether similar non-tree-like signal is found but more pronounced, or whether instead other signals appear. This is, however, beyond the limits of the present article.

3. Multiple tree topologies

The multiple tree topologies method is implemented in the software *BayesPhylogenies* (Pagel & Meade 2004; n.d.). This software provides a Bayesian MCMC (Monte Carlo Markov chain) framework for inferring trees and estimating parameters of evolutionary models for a variety of different data types, including binary lexical data. Bayesian MCMC methods for phylogenetic tree inference are increasingly common for linguistic phylogenetic inference and were used in the analysis of the Austronesian, Indo-European, and Japonic datasets presented above. They have proven so popular because they present a practical solution to finding phylogenetic trees that are well supported by the data (Dunn 2014). The reader can refer to Pagel & Meade (n.d.), Huelsenbeck et al. (2001), Ronquist, van der Mark & Huelsenbeck (2009), Dunn (2014), and Drummond & Bouckaert (2015) for introductions to Bayesian MCMC methods and various software packages that implement them.

In the multiple tree topologies approach, regular MCMC phylogenetic inference is extended using a mixture models approach (Pagel & Meade 2004) in order to estimate two topologies at the same time (Pagel & Meade 2006a). This method can identify different evolutionary histories that may be found in subsets of the dataset. A technical description of the method is not given here; rather, I present an example of the workings of the method using simulated data. Figure 2 presents two random trees, both with 50 language taxa, labeled T1 through T50, in which T50 is the outgroup.

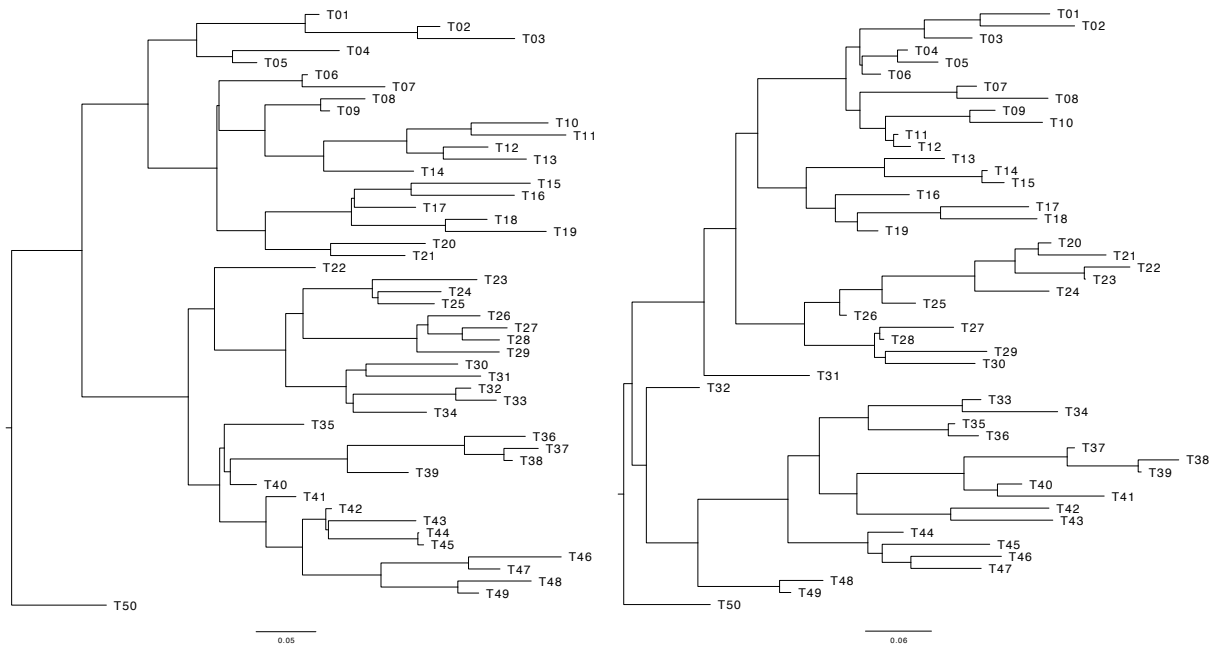


Figure 2. Two random topologies with 50 languages, labeled T1 through T50. Left: tree 1; right: tree 2

The evolution of 10,000 binary traits was simulated across the branches of these two trees using the continuous-time Markov model for discrete trait evolution: 7,500 sites on tree 1 and 2,500 sites on tree 2. After this, phylogenetic tree sets were reconstructed on the basis of these simulated data: once estimating only one topology, and once estimating two topologies. The results are presented in Figures 3 and 4.

Figure 3 presents the maximum clade credibility (MCC) tree of the reconstructed tree set when only a single topology is estimated. It looks remarkably similar to the tree used to simulate 7,500 sites and bears no traces of the groupings found in the tree used to simulate the remaining 2,500 sites. There are a few differences between the trees. One difference is the placement of T17, which in the tree used to generate the data is sister to a group of T15 and T16; in the reconstructed tree, it is sister to a larger (poorly supported) group containing T15, T16, T18, and T19. This is not surprising, given the very short branch leading to T17 in the original tree. T40 is sister to a group containing T36, T37, T38, and T39 in the original tree, but in the reconstructed phylogeny, it groups with T35 – again, because it has a very short branch in the original tree. Other than those two differences, support values are high, and no one would from this single topology guess that 25% of the sites used for this reconstruction have an entirely different evolutionary history from that represented by the maximum clade credibility tree in Figure 3.

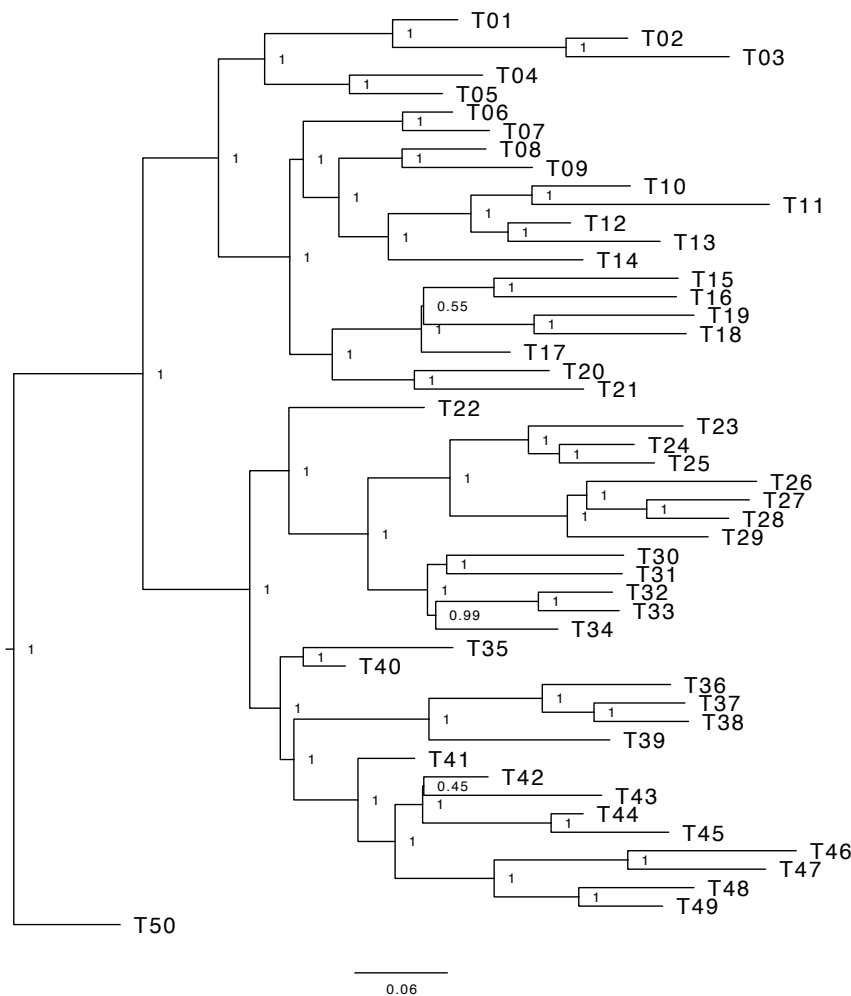


Figure 3. Maximum clade credibility tree of the results of the one topology model on 10,000 simulated binary sites

Figure 4 gives maximum clade credibility trees of the reconstructed tree sets when estimating two topologies. These trees simulate the data with great accuracy: the majority tree has greater support with 76% of the sites, while the minority tree has lower support, 24% of the sites. Not only is the majority tree identical in topology to the original tree that was used for the simulation of the 7,500 sites (T17 and T40 are placed where they were in the original tree), the minority tree captures the topology of the tree used to generate the 2,500 sites with different evolutionary history. Of course, language change will never be as clear-cut as a simulated binary dataset, especially since binary cognate-coded lexical data (the most common type used) is originally multistate and only made binary for the purpose of phylogenetic inference. Likewise, phylogenetic inference models are at best very rough approximations of reality, and would be so even if our data were as noise-free and well-sampled as simulated data. A better simulation would perhaps be to let multistate characters evolve up the branches of the tree, then convert the simulated data to binary before using it for reconstructions (see Pagel & Meade 2006b: 175 and Chousou-Polydouri et al. 2016 on multistate versus binary phylogenetic inference). However, one would then have to specify aspects of multistate character evolution (such as the probability of emergence of new states) that would go beyond the purpose of this article. Therefore, the intention of this simulation is simply to show that the multiple topologies method can in principle recover evolutionary histories that would otherwise be overwhelmed by the patterns found in the majority of the dataset. More specific simulations regarding lexical data should be conducted in the future.

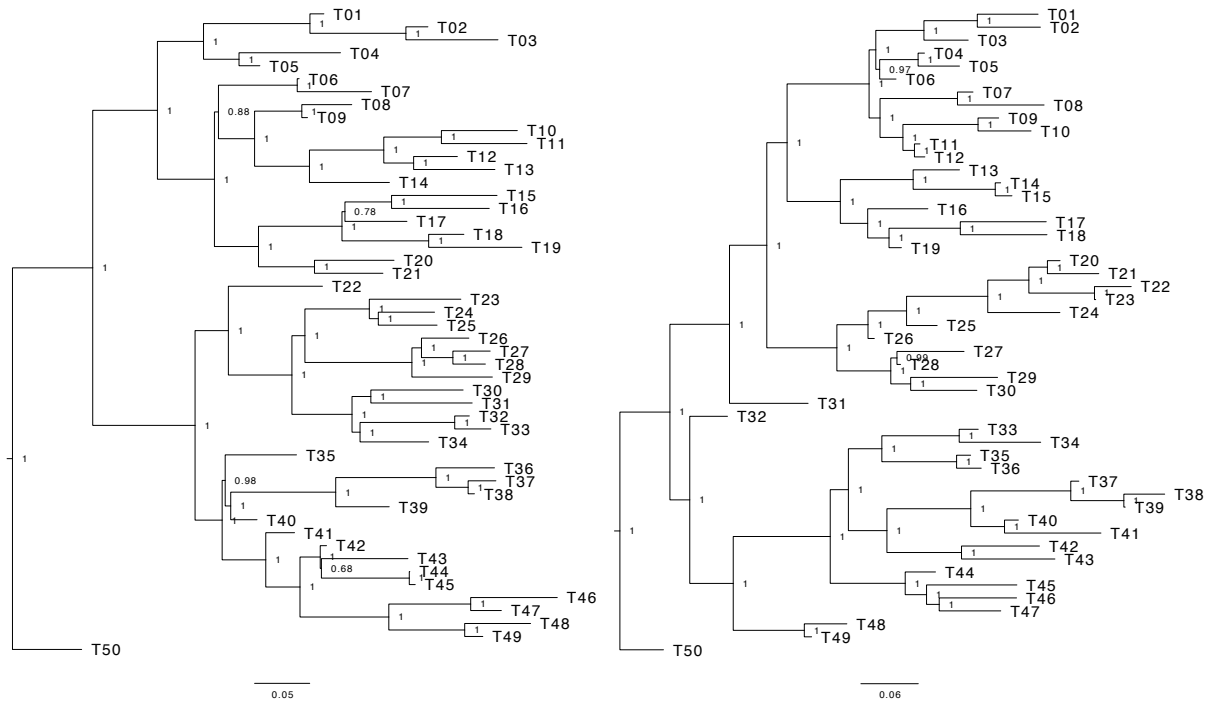


Figure 4. Maximum clade credibility trees of the results of the two-topology model on 10,000 simulated binary sites. Left: The majority tree, supported by 76% of the sites; right: The minority tree, supported by 24% of the sites

There are different patterns that could be identified by the multiple topologies methods. It could be the case that no alternative histories are present, in which case the minority tree will have no highly supported groupings, basically returning “noise.” Another possibility is that there might be an alternative grouping of only a few languages, so only a few highly supported groupings would be found in the minority tree while the rest of the tree would have low support, once again representing noise. One might find that the minority topology picks up on differences in rate, in which case the two tree topologies would be similar but have different branch lengths. All of these different outcomes can be caused by different evolutionary processes, making interpretation potentially difficult. Given that the current study is the first application of the multiple topologies method on linguistic data, all findings must be considered tentative; indeed, in many cases, it is not possible to pinpoint the cause underlying the non-tree-like signal that this method recovers.

It is possible to use the multiple topologies method to assess the log-likelihood of each site (cognate set) for each tree in the posterior sample of both the majority and the minority tree set. What it means to be the majority tree is to have higher log-likelihoods than the minority tree has for at least half of the sites. This functionality is used below to identify cognate sets that are associated specifically with the minority and majority tree and to see how they support particular language groupings.

Before describing the behavior of the multiple tree topologies method in analyzing the four lexical datasets examined in this article, a few technical details need to be mentioned. Eight models of evolution were tested (for further information see the BayesPhylogenies manual, Pagel & Meade n.d.; for an explanation of covarion and gamma rate heterogeneity see Ronquist, van der Mark & Huelsenbeck 2009):

1. standard two-state model (standard)
2. two topologies (top2)
3. covarion (cov)
4. covarion + two topologies (cov+top2)
5. gamma rate heterogeneity (gam)
6. gamma rate heterogeneity + two topologies (gam+top2)
7. covarion + gamma rate heterogeneity (cov+gam)
8. covarion + gamma rate heterogeneity + two topologies (cov+gam+top2)
- 9.

Perhaps comparing these eight models is not necessary – gamma rate heterogeneity is a model well-suited for genetic data, but it normally does not improve the fit of phylogenetic analysis of linguistic data. Phylogenetic analysis of linguistic data typically performs well with the covarion model (see Gray, Drummond & Greenhill 2009; Lee & Hasegawa 2011; Bowerman & Atkinson 2012, and Grollemund et al. 2015 for examples of phylogenetic analyses of lexical datasets in which a covarion model outperforms a gamma rate heterogeneity model; see Chang et al. 2015: 216 for general comments). However, in order to see the interaction between these three different components, all combinations of yes/no covarion, yes/no gamma rate heterogeneity, and yes/no two topologies were investigated. Each model was run for 120 million iterations with a 20 million burn-in; every 100,000th iteration was sampled, creating posterior tree samples of 1,000 trees. Five duplicate analyses were carried out for each model to ensure convergence. No clock, calibration points, or outgroups were provided in order to make the analyses comparable between language families. Nor were trees forced to be ultrametric, i.e. to have the same root-to-tip distance for all tips, as this would be inappropriate for the Indo-European and Japonic datasets, which include non-synchronous language varieties. The result, then, is tree samples in which branch length reflects amount of lexical change rather than chronological time. Appropriate sampling was assessed by measuring the Pearson correlation between iteration and log-likelihood, which should be close to zero. Convergence of each chain was assessed using a linear regression of iteration and log-likelihood, where the β coefficient should be as small as possible and not show a significant downward or upward trend. In Section 4, the converged run of each model with the highest marginal log-likelihood as estimated by stepping stone analysis (Xie et al. 2010) that was conducted after each run is given. The input files for each analysis are shared in the supplementary information files, available on Zenodo ([zenodo.org /record/2653209](https://zenodo.org/record/2653209)).

4. Results

All result files have been made available as supplementary information, found on Zenodo (zenodo.org/record/2653209).

4.1 Austronesian

Table 1 gives an overview of the results of the different models of evolution for Gray, Drummond & Greenhill's (2009) lexical dataset drawn from 400 Austronesian languages. The models are ordered by rank in the second column; the first row details the best-supported model, the last the worst-performing model. Model performance was assessed using marginal log-likelihoods as estimated by stepping stone analysis (Xie et al. 2010). The third column gives the

difference between models in log-likelihood units. The fourth and fifth columns contain the weight associated with each topology for those models in which two topologies were estimated. This represents the number of cognate sets associated with each topology.

Table 1 shows that the cov + gam + top2 model is the best-performing model in terms of marginal log-likelihood. However, it is not altogether clear why this is the case: the weight associated with the minority tree, i.e. the least supported tree, is very low (0.005), so it is unclear what having two topologies adds to the cov + gam model, which scores third. The majority topologies have several well-supported clades and capture the well-known structure of the Austronesian family – these are not considered here in detail due to lack of space, as the current interest is finding evidence for non-tree-like signal. The minority topology is assessed by looking at clades supported at a score of over .5; these clades represent language groupings that are present in at least 50% of the 1000-tree sample. The only analysis for which internal nodes in the minority tree have a posterior support larger than 1% is for the model with two topologies (no covarion, no gamma rate heterogeneity), in which some well-supported internal nodes of pairs of languages are found in the minority topology. This is probably due to the minority topology capturing some aspect of lexical change that is normally captured by the covarion or gamma rate heterogeneity models.

Table 1. Model comparison for Gray, Drummond & Greenhill’s (2009) lexical dataset on 400 Austronesian languages

model	marginal l-lh	marginal l-lh difference	weight majority tree	weight minority tree
cov + gam + top2	-280975		0.995	0.005
gam + top2	-281462	487	0.995	0.005
cov + gam	-281544	82	-	-
gam	-281949	405	-	-
cov + top2	-282265	316	0.994	0.006
cov	-282840	575	-	-
top2	-285127	2122	0.98	0.02
standard	-286478	1516	-	-

a. Marginal l-lh = highest marginal log-likelihood out of 5 runs

b. Marginal l-lh difference = difference in log-likelihood units with better performing model in above row

c. Weight majority tree & weight minority tree = for those models that estimate two topologies, the mean weight associated with the best supported and least supported tree, respectively

The low weightings for the minority trees across all models, plus the lack of well-supported clades in the minority trees for the top-ranking models, make it clear that Gray, Drummond & Greenhill’s (2009) Austronesian dataset does not show any significant support for a second topology. This is highly surprising, as it is well-known that the Polynesian languages likely originated from a dialect chain break-up (Gray, Bryant & Greenhill 2010 and references therein; Kelly & Nicholls 2017). Taking Eastern Polynesian as an example, a language group which was also examined by Gray, Bryant & Greenhill (2010: 3926) and Kelly & Nicholls (2017: 12), we might expect to find Marquesic and Tahitic groupings in one of the two trees, with the other tree reflecting the affinity between Hawaiian or Mangareva with Tahitic; we might also expect to find Rapanui with Marquesan, Mangareva, and Penrhyn. So while we would not see a reconstruction of the dialect chain itself, the multiple topologies method would show divergent signals in the different trees.

4.2 Indo-European

As stated in Section 2, I used a different, updated version of Bouckaert et al.’s (2012) Indo-European lexical dataset kindly provided by Michael Dunn. This dataset has updated cognate coding, and it incorporates 116 rather than 103 languages. The languages present in addition to those in the original dataset are Proto-Indo-European, Albanian_T, Ossetic, Greek_K, Greek_D, Greek_Md, Khaskura, Panjabi_ST, Lithuanian_O, Afrikaans, Pennsylvania_Dutch, Brazilian, and Irish_B. These mostly represent variants of languages otherwise included, many of which have their origin ultimately in the database by Dyen, Kruskal & Black (1992). I will refer to this dataset as “Bouckaert et al.’s (2012) updated lexical dataset.”²

Because it proved difficult to get some of the models to converge, we performed sets of 50+ shorter runs for some models. For the cov + gam + top2 and the cov + top2 models, we used 30 million iterations and a 10 million iteration burn-in, sampling every 20,000th iteration. All other models were run (as stated above) with 120 million iterations and a 20 million iteration burn-in, sampling every 100,000th iteration.

Table 2 gives an overview of the results of the different models of evolution for Bouckaert et al.’s (2012) updated lexical dataset of 116 Indo-European languages. The results for Indo-European look very different from the Austronesian results. The cov + gam + top2 model performs best. In all models that estimate two topologies, the minority topology is quite well supported, with support from 23% to 40% of the sites (cognate sets). Bouckaert et al.’s (2012) maximum clade credibility tree (S1) is reproduced as Figure 5. To first compare Bouckaert et al.’s (2012) results to a model without two topologies, the maximum clade credibility tree of the cov + gam model (the best-supported model without two topologies) is included in Figure 6. What follows are the maximum clade credibility trees of both the minority and majority tree sets for the best-performing cov + gam + top2 model in Figures 7 and 8.

Table 2. Model comparison for Bouckaert et al.’s (2012) updated lexical dataset on 116 Indo-European languages

model	marginal l-lh	marginal l-lh difference	weight majority tree	weight minority tree
cov + gam + top2	-51167		0.60	0.40
cov + top2	-51182	15	0.65	0.35
gam + top2	-51375	193	0.77	0.23
cov + gam	-51450	75	-	-
cov	-51524	74	-	-
gam	-51791	267	-	-
top2	-51832	41	0.64	0.36
standard	-52812	980	-	-

a. Marginal l-lh = highest marginal log-likelihood out of 5 runs

b. Marginal l-lh difference = difference in log-likelihood units with better performing model in above row

c. Weight majority tree & weight minority tree = for those models that estimate two topologies, the mean weight associated with the best supported and least supported tree, respectively

To compare an analysis without two topologies to the maximum clade credibility tree presented in Bouckaert et al. (2012), we can compare Figure 5 and Figure 6. In this comparison, we see the following:

² Some names in the updated dataset are different. They are as follows: Persian_List = Persian; Nepali_List = Nepali; Romani = Gypsy_Gk; Riksmal = Norwegian; English_ST = English; German_ST = German; Romanian_List = Rumanian_List.

1. Higher order subgrouping. Starting from the top, both trees have a Celtic-Romance clade, which is sister to Germanic. The Bouckaert et al. (2012) tree topology then features Balto-Slavic, Indo-Iranian, and the remaining clades in a stepwise fashion. The cov + gam MCC tree has a main split between Celtic-Romance-Germanic and the other subgroups, but this lower major clade has low support for its main splits, e.g. Balto-Slavic and Indo-Iranian (0.32), Balto-Slavic+Indo-Iranian and Albanian-Greek (0.24).
2. Indo-Iranian. The position of Vedic Sanskrit, Old Persian, and Avestan in the cov + gam model deviates from the position of these languages in the Bouckaert et al. (2012) tree. Old Persian and Avestan are sisters and together are a clade sister to Vedic Sanskrit. Wakhi (an Eastern Iranian language) is not in the right place. The Indo-Aryan clades in both trees are similar with minor differences the positions of smaller groups. Interesting is the position of Romani, called Gypsy_Gk in the updated dataset, as an outlier in Bouckaert et al. (2012) but as sister to Kashmiri in the cov + gam tree.
3. Slavic. Polish is not placed correctly (with other West Slavic languages Czech and Slovak) in either tree. Old Church Slavonic is pulled outside of the contemporary Slavic group in the cov + gam tree, while it is correctly placed in the South Slavic group in Bouckaert et al. (2012).
4. Germanic. In the cov + gam tree, Old High German is placed as a sister to West Germanic rather than as a sister to German and Luxembourgish.
5. Romance. The two analyses have the same smaller subgroups within this clade, but their grouping is different. In the cov + gam tree, Umbrian and Oscan are sisters to Latin rather than being placed outside of the clade containing Latin and the modern Romance languages. Sardinian is not a sister to the Italian subgroup.
6. Remaining smaller groups. Tocharian is sister to the Anatolian languages in the cov + gam tree, rather than to the Armenian languages in Bouckaert et al. (2012). The Armenian languages show a rather well-supported affiliation with the Hellenic languages instead.

Note that some of these changes may have been affected by aspects of the phylogenetic analysis by Bouckaert et al. (2012) that were not included in the current analyses, including calibration points on language divergence times and the inference of ancestral geographical locations.

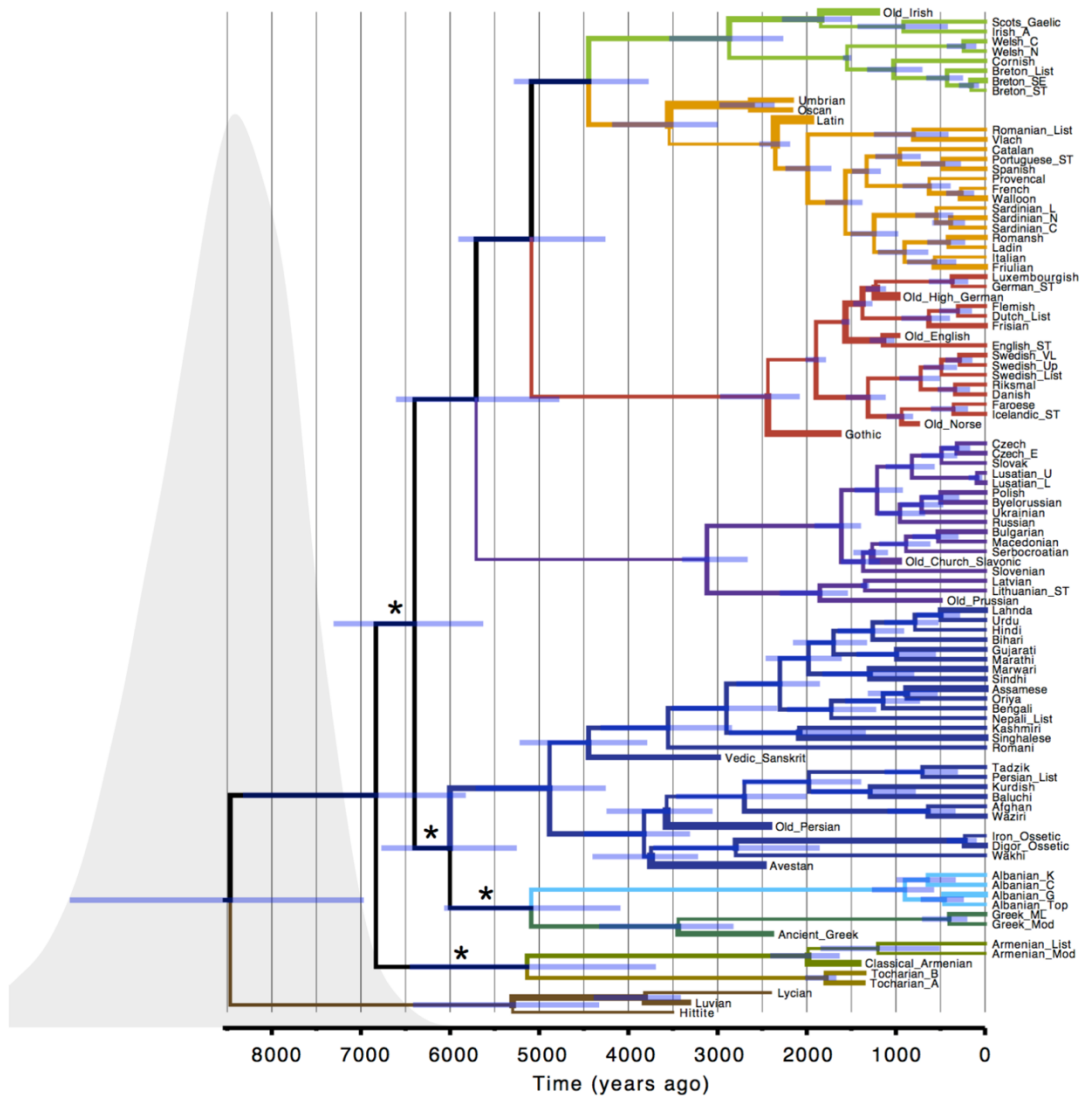


Figure 5. Bouckaert et al.'s (2012) dated maximum clade credibility tree for 103 Indo-European languages. From Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, & Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337.957–960. Reprinted with permission from AAAS

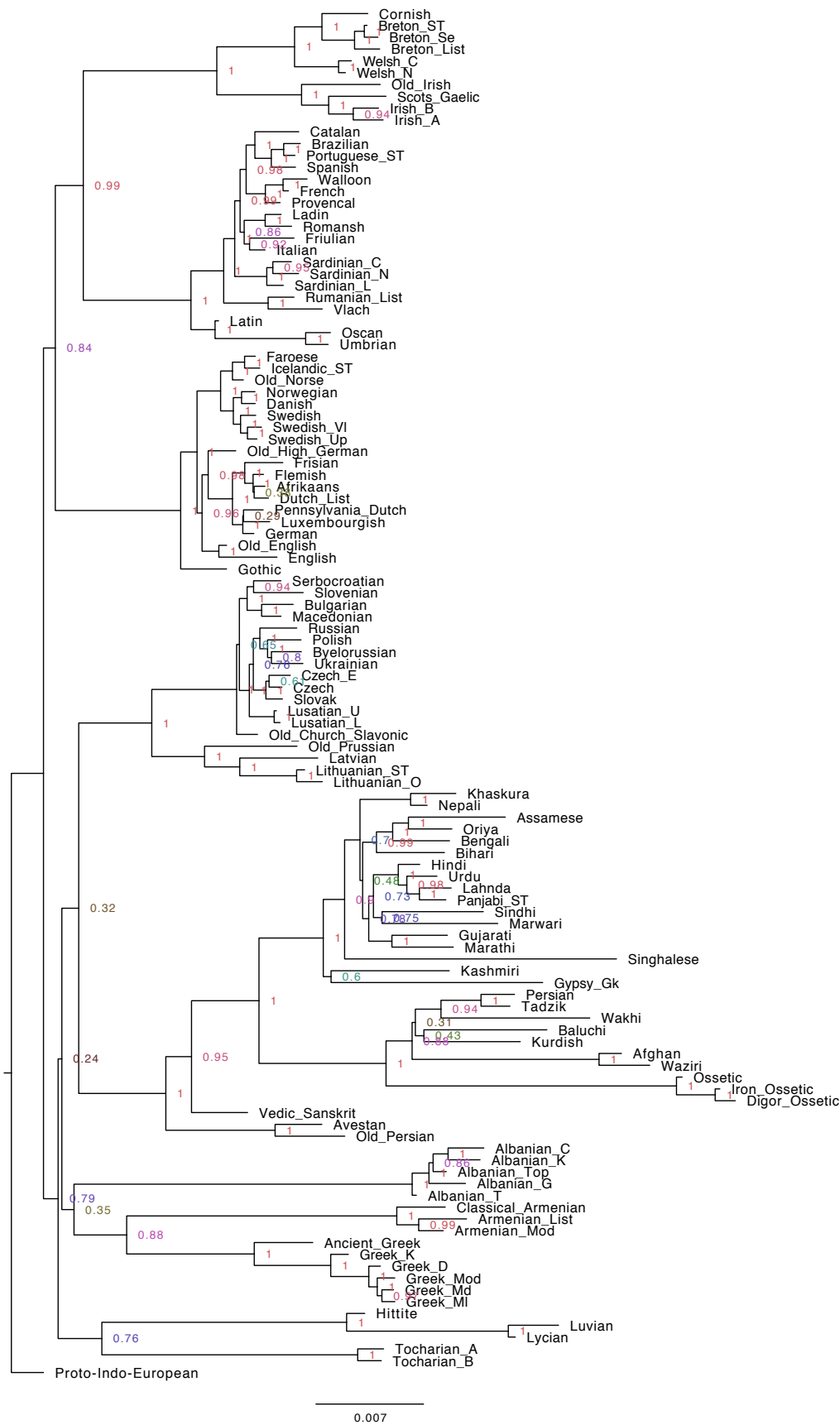


Figure 6. The maximum clade credibility tree for the cov + gam model with the highest marginal log-likelihood from Bouckaert et al. (2012)

The following is a list of points regarding the minority and majority trees of the cov + gam + top2 model, starting at the top (Celtic) and ending at the bottom (Anatolian), comparing them to Bouckaert et al.'s (2012) maximum clade credibility tree in Figure 5.

1. Higher order subgrouping. Neither the minority or the majority tree has the same subgrouping of major branches as the Bouckaert et al. (2012) maximum clade credibility tree has; this is especially true regarding the placement of non-contemporary languages. In the minority tree, Luvian, Lycian, Umbrian, and Oscan form a clade that is sister to Celtic. This tree also deviates by pulling Germanic outside of the Celtic-Romance-Germanic group, albeit with very low posterior probability (0.20). Albanian is placed as sister to Romance. In the majority tree, Oscan and Umbrian are also sister to Celtic, again with very low posterior probability (0.20); this may reflect uncertainty regarding their placement within Celtic-Romance.
2. Celtic. The structure is mostly similar in both the majority and the minority tree and similar to the structure in the Bouckaert et al. (2012) tree. In the majority tree, Cornish and Welsh are sisters, while the minority tree captures the received view that Cornish and Breton are sisters.
3. Romance. In the minority tree, Lycian and Luvian (Anatolian) form a group with Umbrian and Oscan, an odd result with relatively high support: a cursory inspection of the dataset suggests that this is probably caused by a substantial amount of missing data for the same concepts in these four languages. In the minority tree, Catalan is pulled away from Portuguese and Spanish. Other than that, its structure is very similar to that of the cov + gam MCC tree in Figure 6. The majority tree does not recover this structure and has very short branch lengths in part.
4. Germanic. The minority tree has the same topology as the cov + gam MCC tree in Figure 6, with the sole exceptions of the position of Old Norse, which is pulled outside of Northern Germanic, and the slightly more outward position of Frisian. The majority tree does not capture the division between Northern and Western Germanic and looks quite messy. Gothic is pulled inside the clade as sister to Old High German.
5. Slavic. In the minority tree, Old Church Slavonic is placed correctly in South Slavic (albeit with very low posterior probability, 0.23), an improvement upon the cov + gam MCC tree in Figure 6. Lusatian is sister to Russian (again with low posterior probability, 0.23). Polish is again not placed correctly (as it should be placed with other West Slavic languages Czech and Slovak). As was also true for Romance and Germanic, the majority tree does not capture the internal structure of the Slavic subgroup well. Russian is pulled outside as a sister to the entire Slavic group, and South Slavic is not a clade. What it captures rightly is Lusatian as sister to Polish.
6. Iranian. Avestan is not placed correctly (as sister to Ossetic) in either tree; it is attracted to Old Persian in the majority tree and outside of the Indo-Iranian clade in the minority tree. The placement of Wakhi is problematic in both trees (most likely because there are no languages that are closely related to Wakhi included in the dataset). Old Persian is sister to (Modern) Persian in the minority tree.
7. Indo-Aryan. The structure of this group in the minority tree is much like a stepladder, with several languages being pulled outwards from their position in Bouckaert et al.'s (2012) MCC tree, including Sindhi and Singhalese. While the position of Vedic Sanskrit is fine in the minority tree, it is pulled to a position as ancestor to the entire Indo-Iranian clade in the majority tree. The grouping structure of the majority tree is closer to Bouckaert et al.'s (2012) MCC tree, although the placement of Nepali and Khaskura seems off.
8. The backbone of the tree. Bouckaert et al.'s (2012) maximum clade credibility tree has a stepladder structure, with Anatolian splitting off first, Armenian-Tocharian second, Indo-Iranian+Albanian-Greek third, Balto-Slavic fourth, Germanic fifth, and Romance-Celtic last. While most of these internal nodes have high posterior

probabilities (0.80 or 1), two nodes have lower posterior probabilities: the node connecting Indo-Iranian and Albanian-Greek (0.46) and the node connecting this group with the Balto-Slavic-Germanic-Romance-Celtic group (0.49). Neither the majority nor the minority have this same structure, although the majority tree is closer to it. The posterior probabilities of the backbone nodes differ from those of Bouckaert et al. (2012), with those of the minority tree being particularly low.

The impression from studying these two maximum clade credibility trees is that: (i) while the minority tree represents the structure of the Romance, Germanic, and Balto-Slavic groups better than the majority tree, it is the other way around for the Indo-Iranian group; (ii) the majority tree captures the higher-order subgrouping better than the minority tree (with the exception of the placement of Oscan and Umbrian), while the minority tree capitalizes on uncertainty with regard to a small set of languages, including Lycian, Luvian, Oscan, Umbrian, Albanian, and Tocharian; and (iii) the majority tree has markedly short branch lengths within Celtic, Romance, Germanic, and Balto-Slavic, suggesting that the minority and majority tree capture different cognate sets changing at different rates.

Now that the differences in tree topologies of the minority and majority tree have been discussed, I can attempt to explain the high support for two topologies in the Bouckaert et al. (2012) updated dataset. This can be helped by an assessment of the cognate sets that support each tree topology. Using the command “sitelh” in *BayesPhylogenies*, the log-likelihood that each cognate set evolved on the majority and the minority tree can be calculated. An additional analysis using the highest-ranking cov + gam + top2 model was conducted adding the “sitelh” command. The results are very similar to those reported in Table 2 and Figures 7 and 8; they have been included in the Appendix for reasons of space.

In total, Bouckaert et al.’s (2012) updated dataset of 116 Indo-European languages includes 6381 cognate sets. Given 1001 trees in the posterior sample, the log-likelihood of, in total, $6381 \times 1001 = 6387381$ cognate sets is assessed during the site log-likelihood analysis. The log-likelihood difference between the minority and majority tree is assessed by subtracting the log-likelihood value for the majority tree from the log-likelihood value for the minority tree, with any positive values implying greater support for the minority tree. In calculating the differences, we find that almost all of the individual cognate sets support the minority tree at times; at other times, they support the majority tree. This variation occurs with different trees of the posterior sample (support for minority tree: 6045 cognate sets; support for majority tree: 5709 cognate sets). The log-likelihood differences range from 0 to 40.0 (median minority tree: 0.51; median majority tree: 0.39). So most of the support log-likelihood differences are small and inconsequential, a fact that is also indicated by the fact that almost all cognate sets support different tree topologies in different trees of the sample. When we look at log-likelihood differences of support of 5 or larger, we find 586 cognates sets that strongly support the minority tree and 97 cognates sets that strongly support the majority tree. These are almost entirely mutually exclusive, i.e. these are the cognate sets that really fit only with the minority tree or only with the majority tree, not with both. It may seem strange that there are more cognates sets strongly supporting the minority tree while it is weighted less strongly overall than the majority tree (see Appendix, support majority tree: 0.55; support minority tree: 0.45): this is because we are looking at the log-likelihood differences of 5 or larger taken by cognate set, not at all log-likelihood scores. So where the minority tree has more cognate sets with strong support, the majority tree is associated with more cognate sets overall, which favor it with small or medium support.

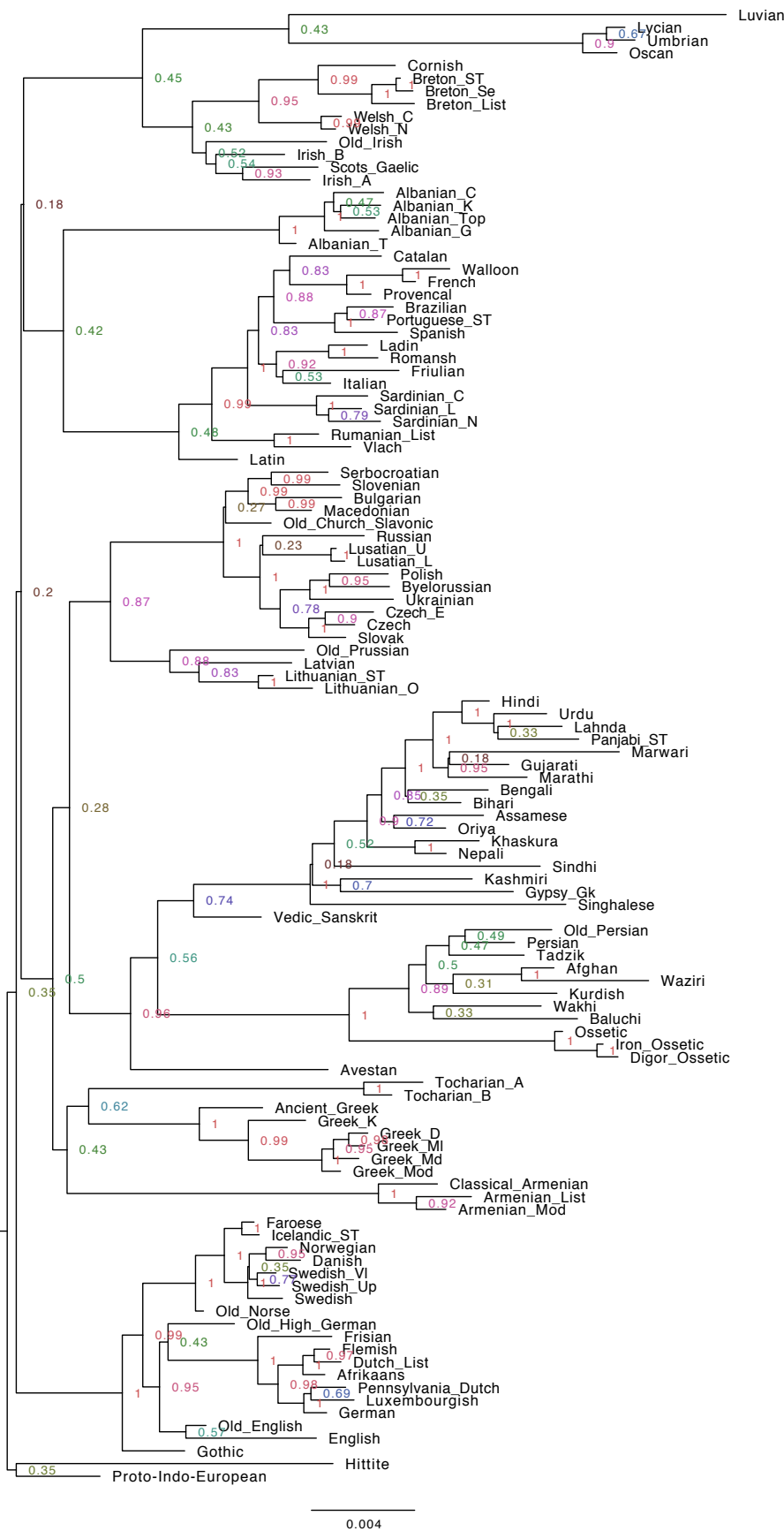


Figure 7. The minority maximum clade credibility tree for the cov + gam + top2 model with the highest marginal log-likelihood (Bouckaert et al. 2012)

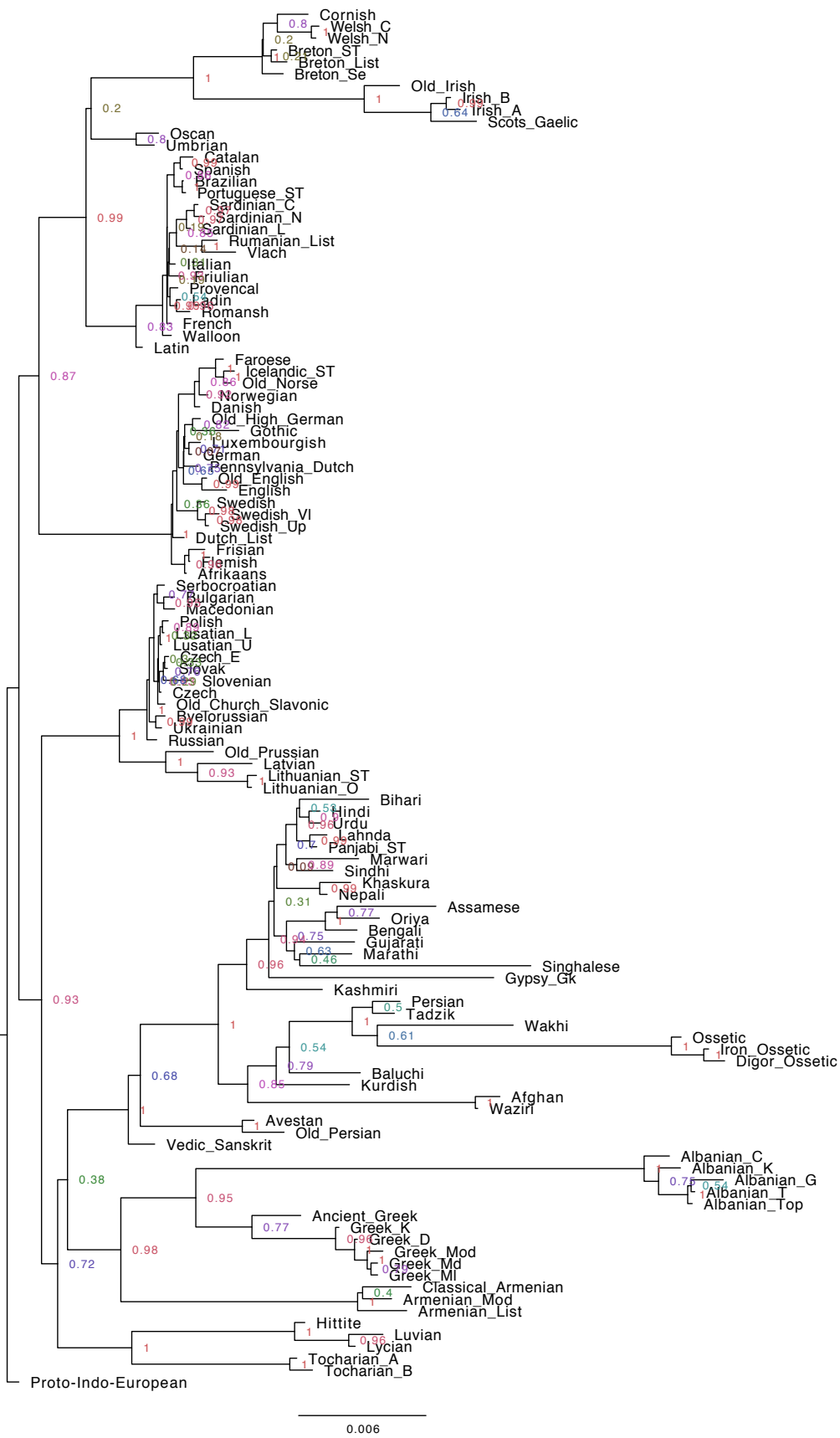


Figure 8. The majority maximum clade credibility tree for the cov + gam + top2 model with the highest marginal log-likelihood (Bouckaert et al. 2012)

To see what kind of topologies emerge from the 586 cognate sets that support the minority tree and the 97 cognate sets that support the majority tree, two additional nexus files were created and two Neighbor Net analyses (Bryant & Moulton 2004, conducted in *SplitsTree*; Huson & Bryant 2006) were conducted. These have been placed in the Appendix for reasons of space (Figures A.3 and A.4). The Neighbor Net analyses show a pattern similar to that of the two topologies analysis, a conclusion we can draw now that we know which cognate sets are responsible for a large part of the two separate signals. Some aspects of the minority and majority tree topology that were remarked upon above are reduplicated:

1. Luvian, Lycian, Umbrian, and Oscan form a clade in the minority Neighbor Net, although this clade is inside Indo-Iranian and not sister to Celtic, as in the minority tree in Figure 7.
2. Catalan seems to be pulled away from Spanish, Portuguese, and Brazilian (Portuguese) in the minority Neighbor Net, just like in the minority tree.
3. The four old Germanic languages (Gothic, Old English, Old High German, and Old Norse) form a clade in the majority Neighbor Net; their attraction is evident in the majority tree from Old High German and Gothic forming a (badly supported) group. In the minority Neighbor Net, Gothic, Old English, and Old High German form a group.
4. The placement of Old Church Slavonic is strange in both the minority Neighbor Net (in a group with Classical Armenian, Pennsylvania Dutch, and Waziri) and the majority Neighbor Net (in a group with Classical Armenian and Pennsylvania Dutch).
5. Four other old or ancient languages – Proto-Indo-European, Vedic Sanskrit, Avestan, and Old Persian – form a clade in the majority Neighbor Net; their attraction is evident in the majority tree based upon the fact that Avestan and Old Persian form a group with Vedic Sanskrit just outside on the next level of a ladder-like structure. Proto-Indo-European, Vedic Sanskrit, and Avestan also form a group in the minority Neighbor Net.

The cognate sets associated most strongly with the minority and majority trees generate Neighbor Nets in which some of the old and ancient languages mentioned above cluster together. The reason for this could potentially be methodological. Cognate coding is usually done with the help of etymological dictionaries and other materials that discuss the history of words. Etymological dictionaries of ancient languages often focus on relationships of words with other ancient languages: i.e., in an etymological dictionary of Gothic, we would expect to find whether a given word has a reflex in Greek, Latin, Old English, Old Norse, or Old Irish, for instance. Contemporary languages, especially non-Germanic ones, would not be featured in this dictionary (except for aberrations as early borrowings into Romance, etc.). By using material from different etymological dictionaries, it is usually possible to find the relevant word histories for both ancient and contemporary languages; nevertheless, the focus on non-contemporary languages in such dictionaries may produce a greater number of cognates between old or ancient languages.

The Neighbor Nets also reduplicate the difference in branch length found in the minority and majority trees. The minority Neighbor Net has long branch length for Germanic, Romance, and Slavic, while the majority Neighbor Net has much shorter branches for these groups, more in line with typical branch length found in the rest of the network. The branch length differences are very striking when comparing the minority and majority trees and can be related to differences in the rate of evolution of the cognate sets most strongly associated with them: the longer branches of the minority tree relate a greater amount of lexical change. Bouckaert et al.'s (2012) updated dataset includes 207 concepts. The 97 cognate sets that show high support for the majority tree feature 79 unique concepts, while the 586 cognate sets with high support for the minority tree feature 163. The concepts associated with the minority tree have significantly

higher rates of lexical replacement (mean 3.49) than those associated with the majority tree (mean 2.85) (t-test, $t = -2.65$, $p > 0.01$; rates of lexical change taken from Pagel, Atkinson & Meade 2007). This suggests that the minority topology, in addition to showing differences in language grouping as discussed above, also captures faster-changing concepts in Germanic, Romance, and Slavic.

The reason for finding support for two topologies that are similar in terms of major subfamily grouping but have differences in branch length and topology in Bouckaert et al.'s (2012) updated lexical dataset must, then, be primarily that there is a significantly large group of cognate sets in Germanic, Romance, and Slavic that change at markedly higher rates of evolution. The differences in topology are at least partly indirectly caused by that – the minority topology captures more detailed information on language relationships within these three groups because the cognate sets associated with it display more change, while the majority tree captures the higher-order subgrouping of the subfamilies better. Close associations between old and ancient languages emerge in both the minority and majority trees and may point to methodological aspects of cognate-coding. The Luvian-Lycian-Umbrian-Oscan clade found in the minority tree is probably caused by missing data for a highly overlapping set of concepts in these four languages, but it serves as a good illustration of the multiple topologies method, i.e. that such a distinctive pattern is picked up on by only one of the trees. For the Indo-European analysis, no clear indication of the role of reticulation is found (despite earlier findings such as Nelson-Sathi et al. 2011 and List et al. 2013). This is not the case for Japonic and Sinitic, as is discussed next.

4.3 Japonic

Table 3 gives an overview of the results of the different models of evolution for Lee & Hasegawa's (2011) lexical dataset for 59 Japonic languages.

The best-scoring model for the Japonic dataset is the cov + gam + top2 model. That model has limited support for two topologies, with the minority tree only getting a 0.04 weight. The other models, however, show more support for the minority topology. The cov + gam + top2 model results are discussed first, then compared to the cov + top2 model, which is the second-best scoring model. The reason for this is that the cov + top2 model has much larger support for the minority tree than the cov + gam + top2 model; it is interesting to see what causes this difference.

A maximum clade credibility tree of both the minority and majority trees of the highest scoring cov + gam + top2 model is included in Figures 10 and 11, respectively. First, I carry out a comparison of the majority tree with Lee & Hasegawa's (2011) tree (their Figure 2, here reproduced as Figure 9). The major split in the Japonic languages is one between the Ryukyuan languages and the mainland Japanese languages. This split is likewise found in the majority tree.

Table 3. Model comparison for Lee & Hasegawa’s (2011) lexical dataset for 59 Japonic languages

model	marginal l-lh	marginal l-lh difference	weight majority tree	weight minority tree
cov + gam + top2	-6759		0.96	0.04
cov + top2	-6783	24	0.81	0.19
cov + gam	-6855	72	-	-
gam + top2	-6886	31	0.76	0.24
cov	-6900	14	-	-
top2	-6999	99	0.83	0.17
gam	-7201	202	-	-
standard	-7583	382	-	-

a. Marginal l-lh = highest marginal log-likelihood out of 5 runs

b. Marginal l-lh difference = difference in log-likelihood units with better performing model in above row

c. Weight majority tree & weight minority tree = for those models that estimate two topologies, the mean weight associated with the best supported and least supported tree, respectively

Aside from that, it is difficult to make a comparison, because several internal nodes in Lee & Hasegawa’s (2011) MCC tree are poorly resolved and the same is true of the current majority tree. The major groupings, however, are similar (from top to bottom): Hokkaido-Yamanashi (although Tokyo is not included in the majority tree), Gifu-Aichi, Aomori-Miyagi-Akita-Yamagata-Iwate-Fukushima-Ibaragi-Tochigi, Shiga-Nara-Kyoto-Osaka-Wakayama (although in Lee & Hasegawa’s (2011) MCC tree, Shiga is not included in this group), Hyogo-Tokushima-Kagawa-Ehime-Kochi-Okayama (although in Lee & Hasegawa’s (2011) MCC tree, this group includes Hiroshima and Shiga), and Fukuoka-Saga-Nagasaki-Kumamoto-Miyazaki-Oita-Kagoshima. The grouping of Chiba-Gunma-Saitama-Kanagawa as found in the majority tree is not found in Lee & Hasegawa’s (2011) MCC tree, and neither is the Nigata-Toyama-Ishikawa-Fukui-Nagano grouping.

In the minority tree, only a few groupings are supported at a reasonably high level: Miyagi-Fukushima-Miyazaki (0.75), Saitama-Tokushima-Nagano (0.43), Toyama-Okayama (0.60), Aomori-Osaka-Shiga-Wakayama (0.38), and Irafu-Ishigaki (0.49). None of these groupings are found in either Lee & Hasegawa’s (2011) MCC tree or in the majority tree.

As was done for Indo-European, additional analyses using the “sitelh” command in *BayesPhylogenies* were conducted to calculate the log-likelihood that each cognate set evolved on the majority and the minority tree and thus to assess the signal retrieved by the minority tree. In Figure 12, an overview is presented of the 83 cognate sets that support the minority tree for the cov + gam + top2 model. For most of these 83 cognate sets, support is rather low: a line has been drawn at 2, a point which is rather arbitrary but which helps the reader see which cognate sets really stand out. The meanings for cognate sets that support the minority tree with a log-likelihood difference of 2 or more are given on top of the bars.

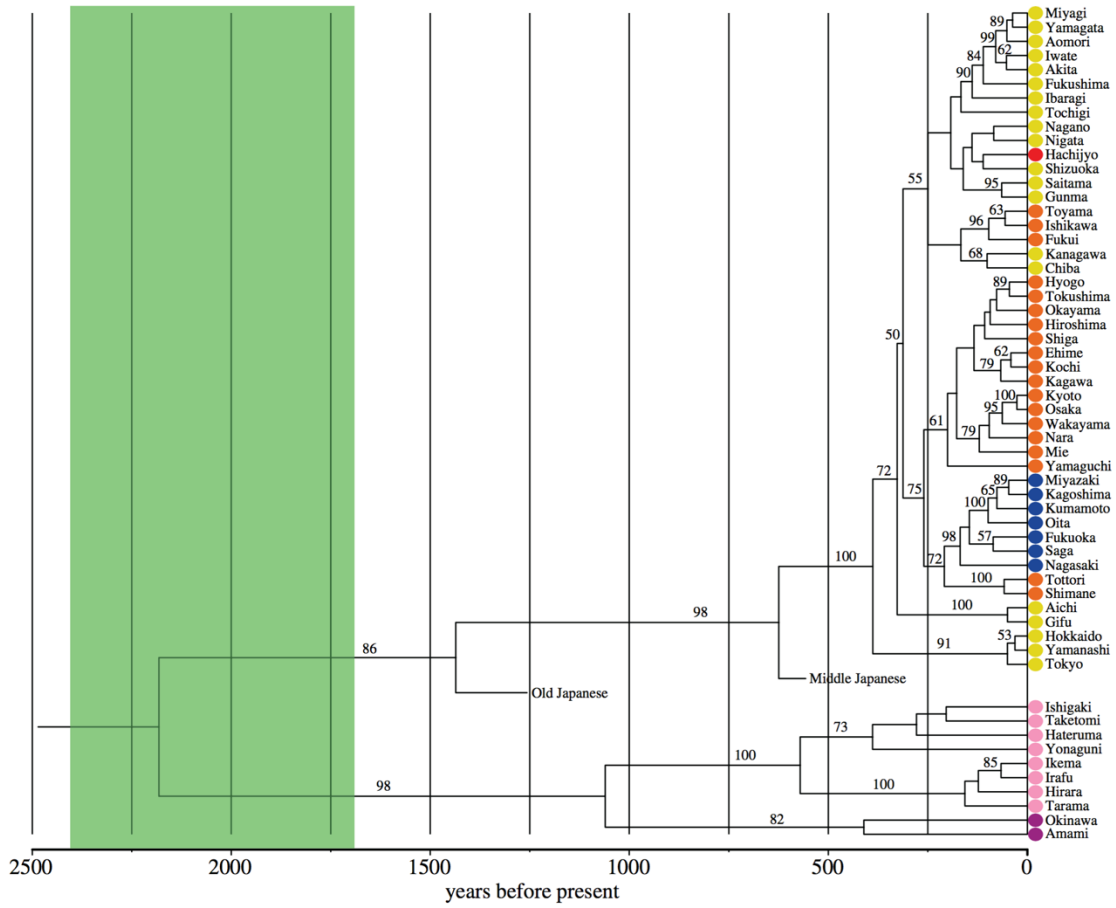


Figure 9. Lee & Hasegawa's (2011) dated maximum clade credibility tree of the Japonic languages. From Lee, Sean, & Toshikazu Hasegawa. 2011. Bayesian Phylogenetic Analysis Supports an Agricultural Origin of Japonic Languages. *Proceedings of the Royal Society B* 278:1725.3662–3669. By permission of the Royal Society.

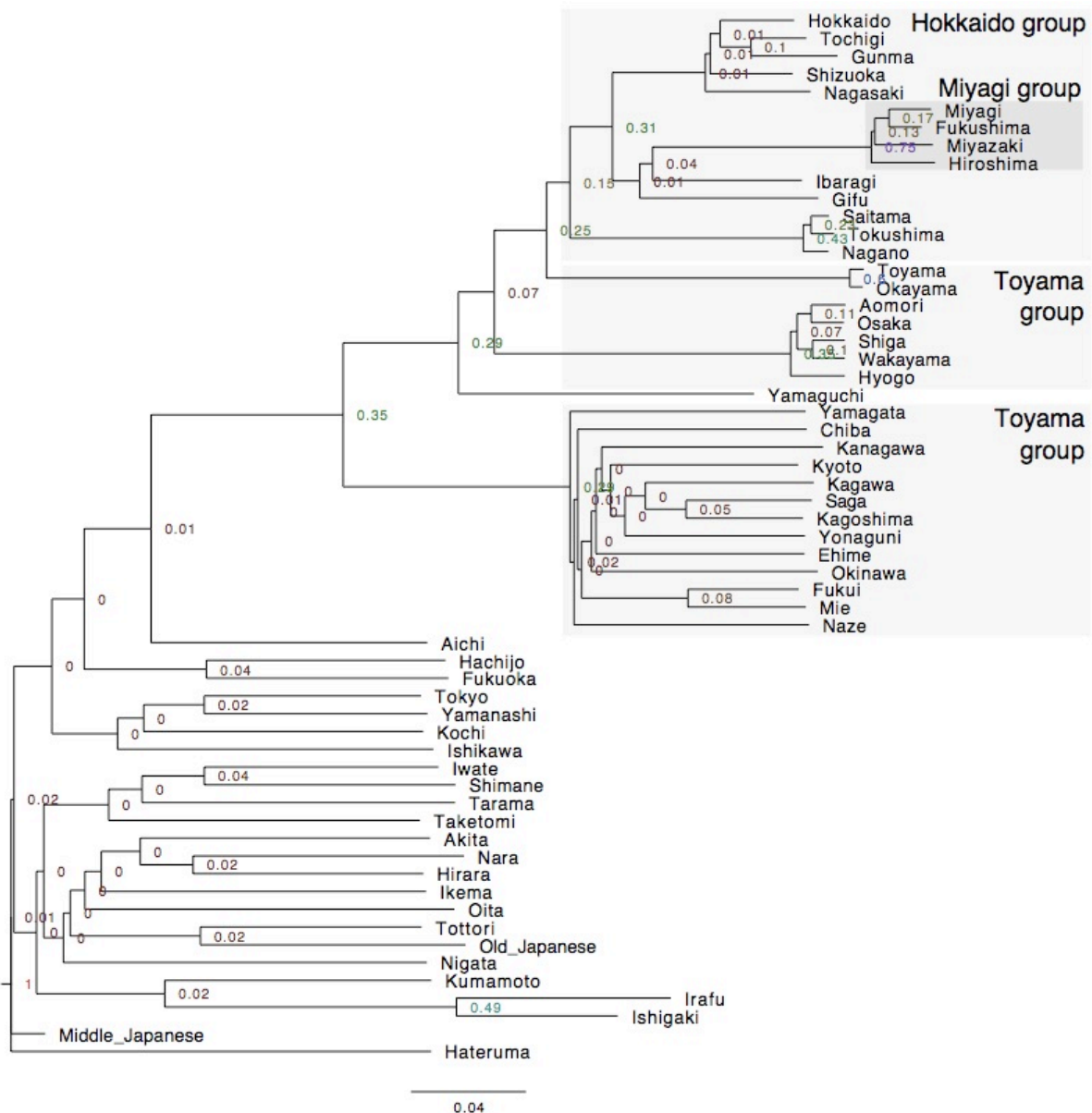


Figure 10. The minority maximum clade credibility tree for the cov + gam + top2 model with the highest marginal log-likelihood (Lee & Hasegawa 2011). Gray boxes represent subgroups with relatively high posterior support that are discussed in text

Three pairs of cognate sets for three concepts stand out: cognate sets for ‘tongue’, ‘to open, unlock’, and ‘night’. These sets support the minority tree with a large log-likelihood difference – over 15 log-likelihood units. It could be the case that these three pairs of cognate sets reflect a subgrouping of languages that is well-supported by the minority tree topology. In these three cases, there are only two cognate sets for each concept; therefore, both cognate sets make the same language grouping twice, which is why both cognate sets for each of these three concepts pattern similarly. Figure 12 will be discussed further below.

The trees in Figures 10 and 11, which give the majority and minority MCC trees for the cov + gam + top2 model and analyze which cognate sets support the minority tree in Figure 12, can be compared to the results from the cov + top2 model, which has far greater support for the minority topology. The trees are included in Figures 13 and 14. The majority topology of this model looks very similar to the majority tree of the cov + gam + top2 model, except for the placement of Nigata and Nagasaki. The support values are similar as well. When comparing

the minority tree topologies (Figures 10 and 13), these are likewise similar, but the reasonably supported groups – for instance, the Miyagi group – have higher posterior support values (0.75 in the cov + gam + top2 model for Miyagi-Fukushima-Miyazaki; 0.90 in the cov + top2 model for Miyagi-Hiroshima-Miyazaki-Fukushima). Thus, the minority topology captures the alternative signal better without gamma rate heterogeneity than it does with gamma rate heterogeneity.

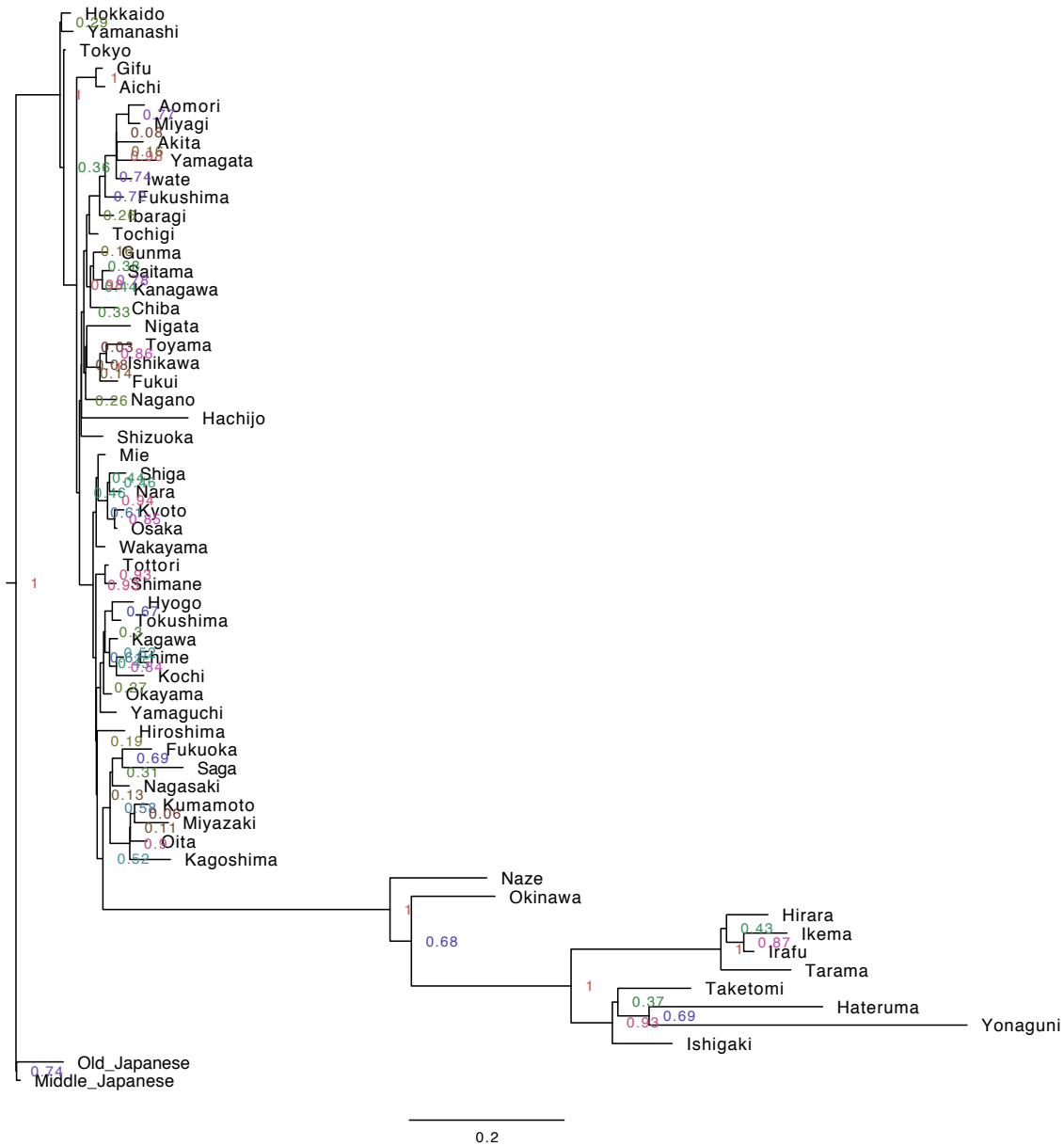


Figure 11. The majority maximum clade credibility tree for the cov + gam + top2 model with the highest marginal log-likelihood (Lee & Hasegawa 2011)

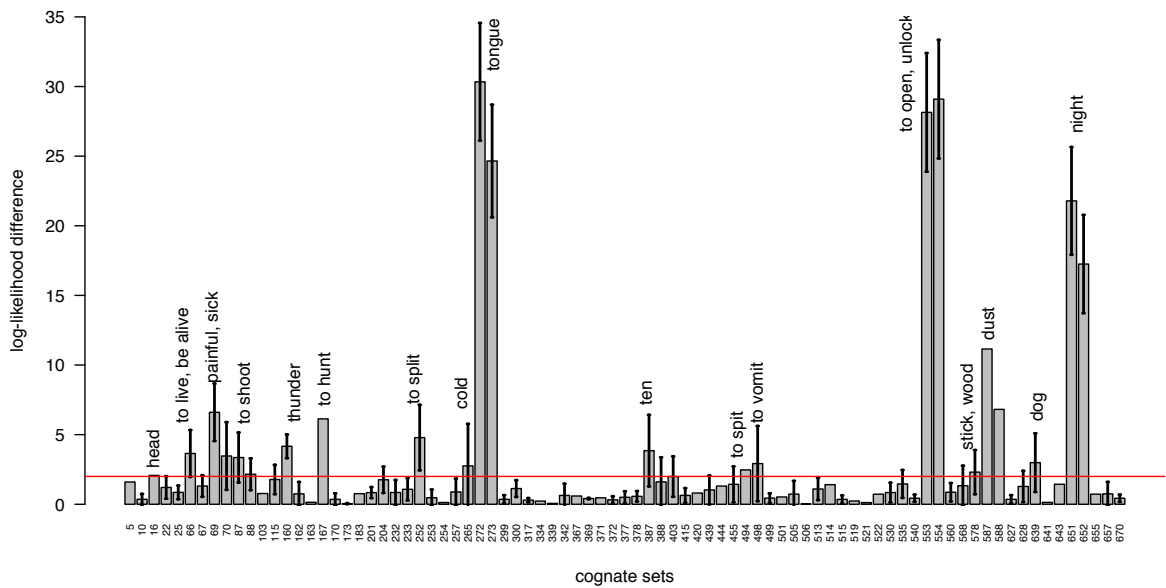


Figure 12. 83 cognate sets (out of 675) that support the minority tree for the cov + gam + top2 model with the highest marginal log-likelihood (Lee & Hasegawa 2011)

The increased support for the minority tree is reflected by Figure 15, which plots the 192 cognate sets that support the minority tree of the cov + top2 model. This is more than double the number of cognate sets associated with the minority tree in the cov + gam + top2 model. Many of the same cognate sets are encountered here: ‘to live, be alive’, ‘painful, sick’, ‘to shoot’, ‘thunder’, ‘to split’, ‘cold’, ‘tongue’, ‘ten’, ‘to vomit’, ‘to open, uncover’, ‘stick, wood’, ‘dust’, ‘night’.

It is easy to see how the three highest scoring cognate sets (‘tongue’, ‘to open, uncover’, ‘night’) inform the structure of the minority topology. In Figure 10, which represents the minority maximum clade credibility tree for the cov + gam + top2 model, the top right-most clade designated as “Hokkaido group” (with posterior probability 0.15) captures the division between the languages made by the cognate sets for ‘tongue’ (the labels “Hokkaido group” and others in Figure 10 and 13 are arbitrary and just for convenience). The ‘to open, uncover’ cognate sets group together Miyagi, Fukushima, Miyazaki, and Hiroshima, a small clade with 0.75 posterior indicated by “Miyagi group,” and a larger group designated as “Toyama group.” Lastly, the ‘night’ cognate sets group together Saitama, Tokushima, Nagano, the upper part of the Toyama group, and Yamaguchi. In Figure 13, which represents the minority maximum clade credibility tree for the cov + top2 model, we find in part the same groupings. There is also a “Hokkaido group,” although Saitama, Tokushima, and Nagano are missing from it; a “Miyagi group,”; and what I called in Figure 13 the “Aomori group,” which again groups together languages on the basis of the ‘night’ cognate sets. Undoubtedly, the other cognate sets with reasonable support for the minority topologies in Figure 10 and 13 have similar (but partial) groupings, leading to the structure observed in the minority tree.

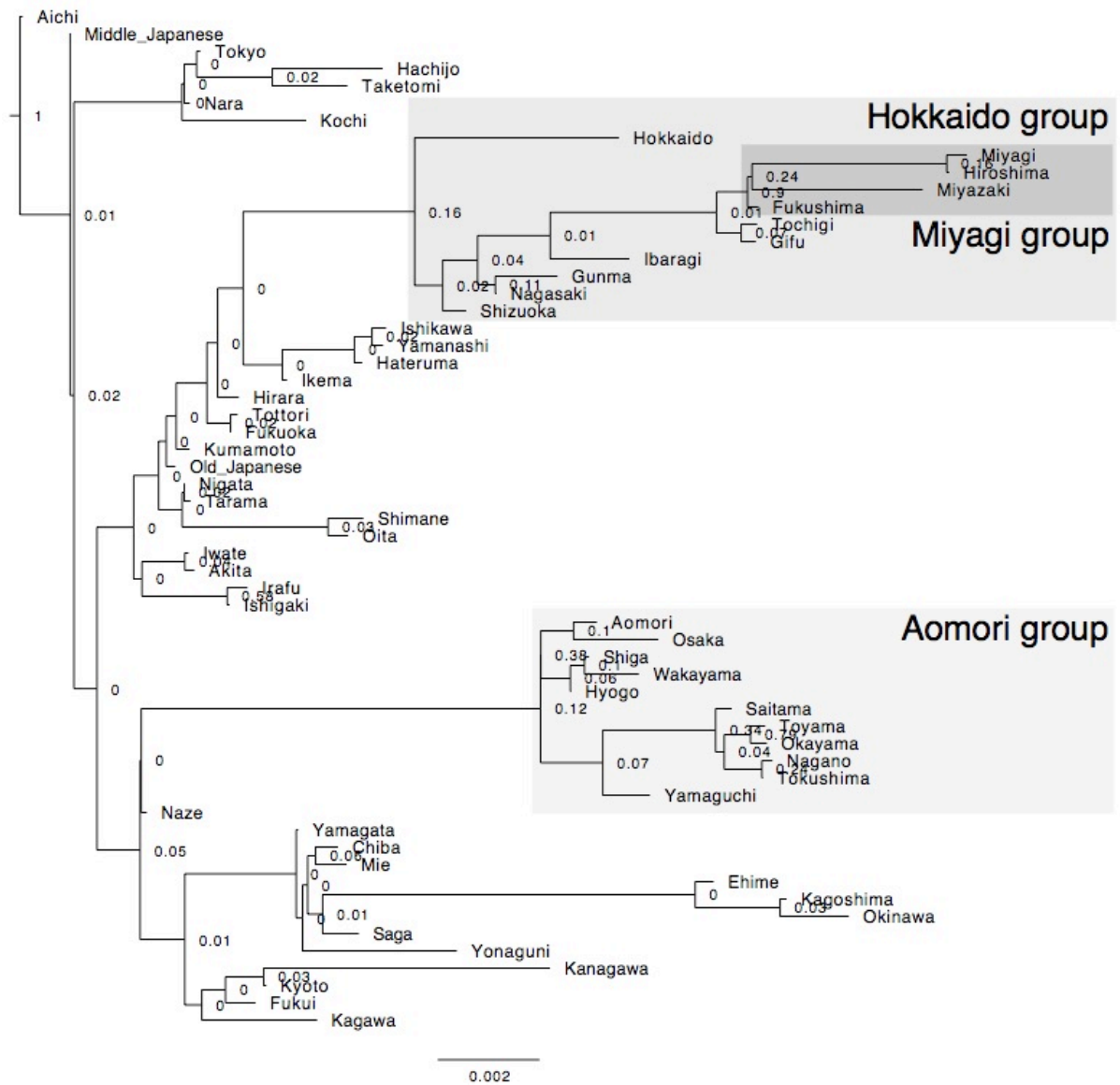


Figure 13. The minority maximum clade credibility tree for the cov + top2 model with the highest marginal log-likelihood (Lee & Hasegawa 2011). Gray boxes represent subgroups with relatively high posterior support that are discussed in text

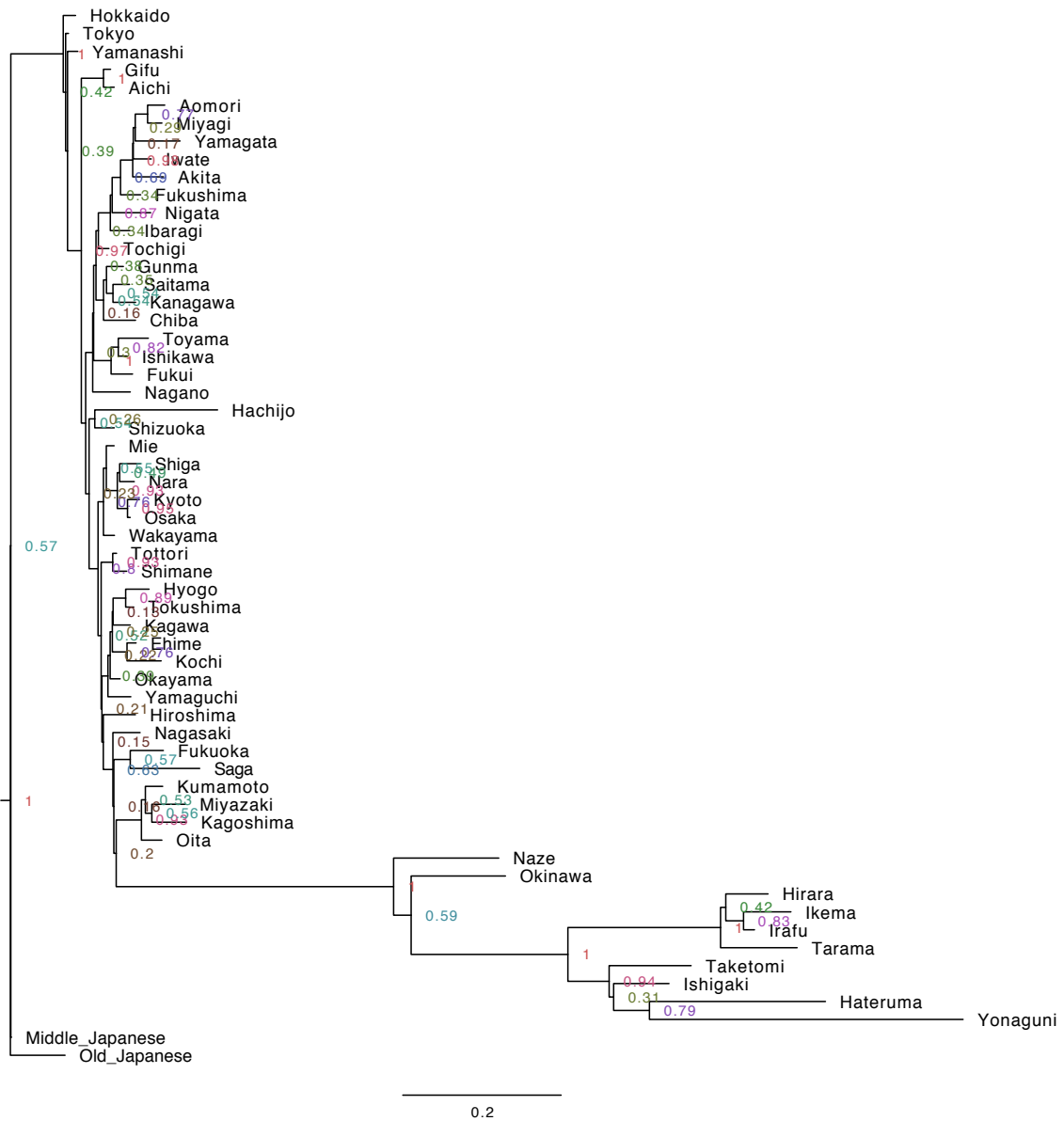


Figure 14. The majority maximum clade credibility tree for the cov + top2 model with the highest marginal log-likelihood (Lee & Hasegawa 2011)

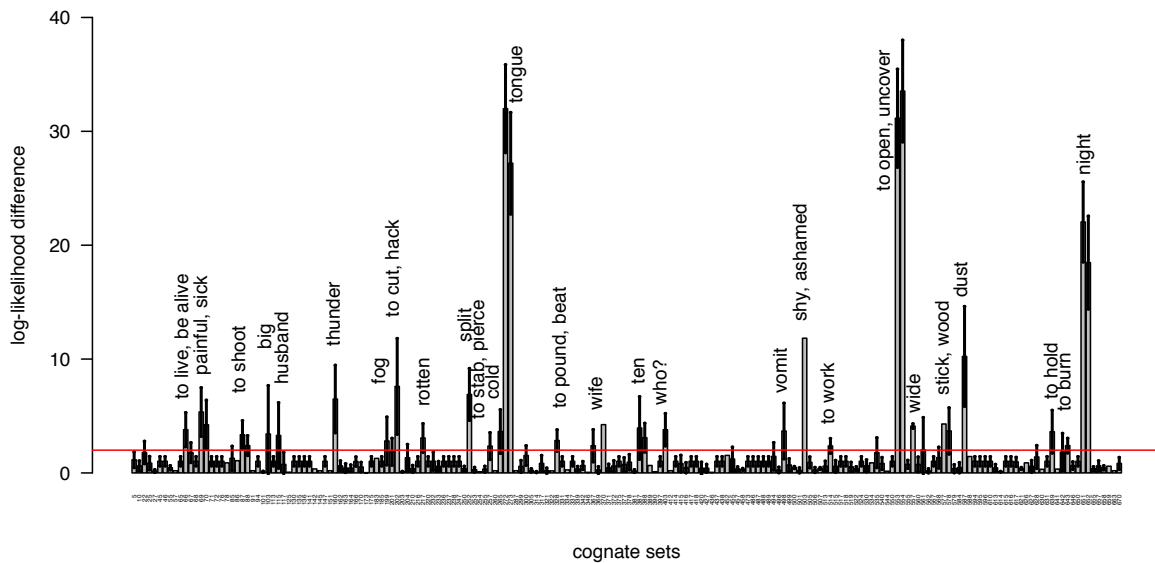


Figure 15. 192 cognate sets (out of 675) that support the minority tree for the cov + top2 model with the highest marginal log-likelihood (Lee & Hasegawa 2011)

4.4 Sinitic

The Sinitic dataset used in this study comes from Wang (2004), a set which excludes Old Chinese. This set was obtained from the Chinese Dialect Database (List 2017) (see above). Table 4 gives an overview of the results of the different models of evolution for Ben Hamed & Wang’s (2006) lexical dataset for 23 Sinitic languages. The best-supported model is the cov + top2 model. However, the marginal log-likelihood difference with the second-best performing model, cov + gam + top2, is not very large (10 log-likelihood units). All models that estimate a second topology find a large amount of support for it. Figure 16 features the majority tree and the minority tree of the best-performing cov + top2 model, while Figure 17 features the same for the second-best model, cov + gam + top2. Comparing the cov + top2 results in Figure 16 to the cov + gam + top2 results in Figure 17, it is striking to see that there are no large differences. The majority trees have the same topology, but the cov + gam + top2 majority tree is better supported (higher posterior values on the internal nodes). Strangely enough, the cov + gam + top2 minority tree is also much better supported than the cov + top2 minority tree. They have (almost) the same structure, but the cov + top2 minority tree has very low posterior values for the internal nodes. The difference between the two models is that the cov + gam + top2 model has an almost-equal division of weighting between the majority and minority tree (0.55 and 0.45, respectively), while the cov + top2 model has more unequal weighting favoring the majority tree (majority tree 0.62, minority tree 0.38).

The majority and minority tree structures for both the cov + top2 model and the cov + gam + top2 model can be compared to network analyses in Ben Hamed & Wang (2006) and List (2015). Ben Hamed & Wang (2006) report the results of a Neighbor Net analysis, which was redone by the current author for the sake of better legibility and is presented in Figure 18. The Neighbor Net analysis retrieves the traditional groupings established for Sinitic languages (Norman 1988:181ff):

- Mandarin: Wuhan, Yingshan, Chengdu, Ningxia, Beijing, Taiyuan, Yuci
- Wu: Ningbo, Shanghai, Shanghai B, Suzhou, Wenzhou
- Gan: Nanchang, Anyi
- Xiang: Changsha, Shuangfeng

- Hakka: Liancheng, Meixian
- Yue: Guangzhou
- Min: Fuzhou, Taibei (Taiwan), Zhangping, Xiamen

Table 4. Model comparison for Ben Hamed & Wang’s (2006) lexical dataset on 23 Sinitic languages

model	marginal l-lh	marginal l-lh difference	weight majority tree	weight minority tree
cov + top2	-9007		0.62	0.38
cov + gam + top2	-9017	10	0.55	0.45
gam + top2	-9046	29	0.62	0.38
cov + gam	-9052	6	-	-
cov	-9060	8	-	-
gam	-9106	46	-	-
top2	-9138	32	0.58	0.42
standard	-9392	254	-	-

- a. Marginal l-lh = highest marginal log-likelihood out of 5 runs
 b. Marginal l-lh difference = difference in log-likelihood units with better performing model in above row
 c. Weight majority tree & weight minority tree = for those models that estimate two topologies, the mean weight associated with the best supported and least supported tree, respectively

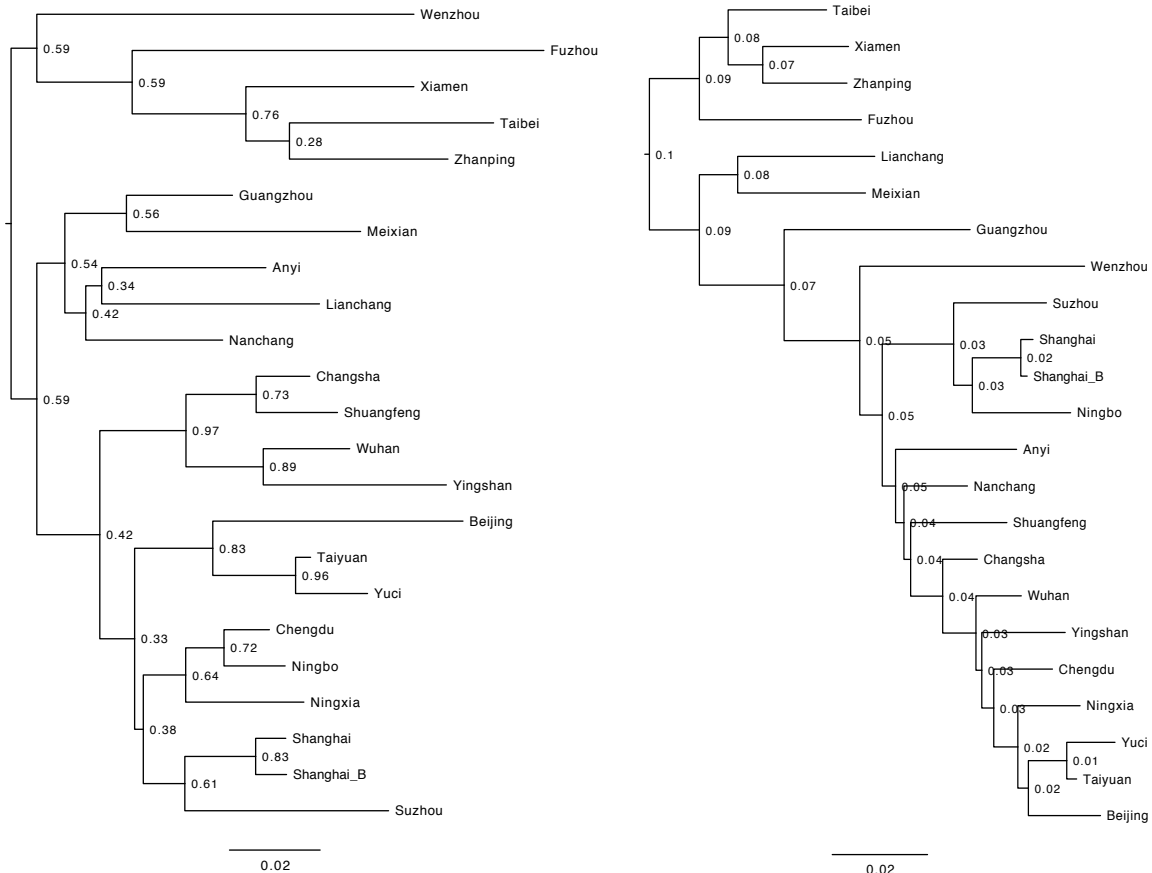


Figure 16. Maximum clade credibility trees for the cov + top2 model for Ben Hamed & Wang’s (2006) Sinitic dataset. On the left is the majority topology (support 0.62); on the right is the minority topology (support 0.38)

Norman (1988:181–183) distinguishes three higher-order subgroups: the Northern/Mandarin group, the Central group consisting of Wu, Gan, and Xiang languages, and the Southern group of Hakka, Yue, and Min languages. Ben Hamed & Wang’s (2006) Neighbor Net reflects the Southern group, but we cannot recognize the Central group as such; rather, the main division is between Northern+Central and Southern.

Since Ben Hamed & Wang’s (2006) Neighbor Net in Figure 18 captures traditional groupings as well as reticulate signals quite well, it is possible to use it as a benchmark for the behavior of the multiple topologies results. To start at the top of the majority topology in Figures 16 and 17, Wenzhou is included in the Min group of Fuzhou-Taibei-Zhangping-Xiamen, rather than with the Wu languages. Next, Guangzhou, the only Yue language, is sister to Meixian, one of the two Hakka languages. The other Hakka language, Liancheng, forms a group with Anyi and Nanchang, the two Gan languages. Further down, the two Xiang languages, Changsha and Shuangfeng, form a group with two of the Mandarin languages, Yingshan and Wuhan. This group splits up the Mandarin languages above and below it: first Beijing, Taiyuan, and Yuci; and lower down, a group of the remaining two Mandarin languages, Ningxia and Chengdu, with Ningbo (Wu). At the bottom, the remaining three Wu languages – Shanghai, Shanghai B, and Suzhou – form a group. All in all, the majority tree does not match Ben Hamed & Wang’s (2006) Neighbor Net very closely at all: most clades contain languages from 2 different language groups.

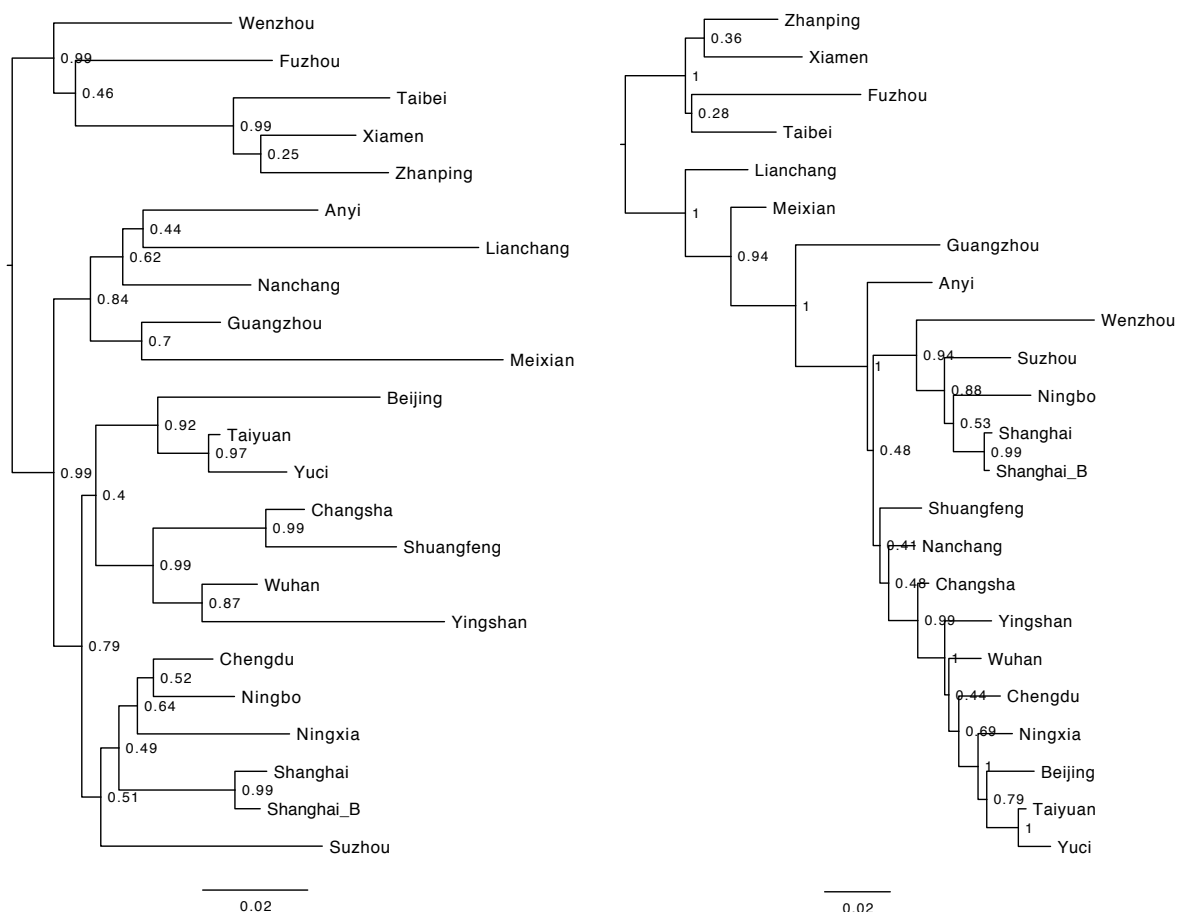


Figure 17. Two topologies for the cov + gam + top2 model for the Sinitic dataset (Ben Hamed & Wang 2006). On the left is the majority topology (support 0.55); on the right is the minority topology (support 0.45)

Next, let us consider List (2015). List (2015) constructs minimal lateral networks on the basis of three different reference trees: (i) Laurent Sagart’s *Arbre des Dialectes Chinois*, (ii) Jerry Norman’s *Southern Chinese Hypothesis*, and (iii) Yóu Rǔjié’s *Hànyǔ Fāngyán Shùxíngtú* “Tree chart of Chinese dialects.” These trees are abbreviated by List (2015) as *Arbre*, *Southern Chinese*, and *Shùxíngtú*, respectively, a naming system which will be followed here for the sake of convenience. List (2015:36–37) additionally includes three reference trees that were reconstructed using distance-based methods and maximum parsimony, but since List (2015) emphasizes the need for the “correct” reference tree for minimal lateral networks to perform optimally, these additional trees will not be considered here.

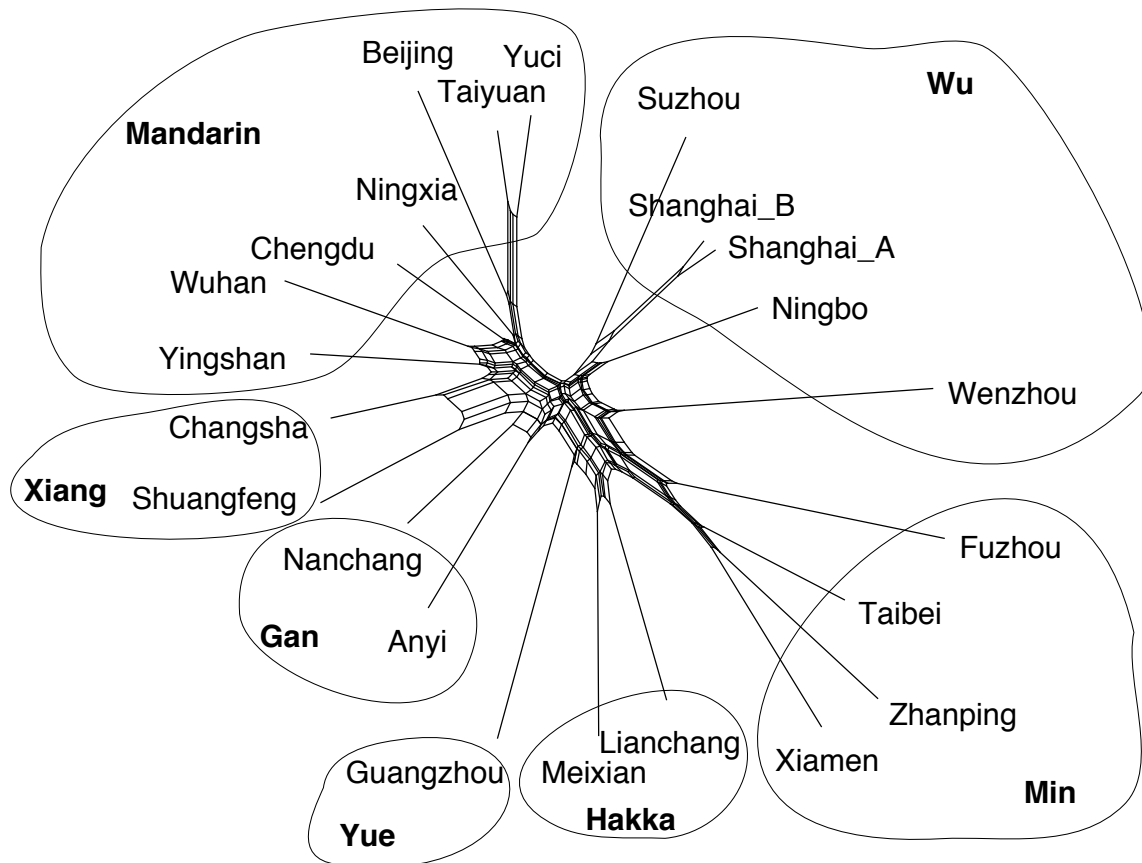


Figure 18. A Neighbor Net produced with Ben Hamed & Wang’s (2006) dataset. It is identical to their Neighbor Net (their Figure 10, Ben Hamed & Wang 2006: 54), except for the inclusion of Old Chinese

The minimal lateral network based on the *Arbre* (reproduced in Figure 19) and *Southern Chinese* reference trees distinguishes the following clades (List 2015, Supplementary Information IV, going from right to left):

1. Zhangping, Xiamen, Taipei, Fuzhou (Norman’s 1988 Min group)
2. Liancheng, Meixian, Guangzhou (Norman’s 1988 Hakka + Yue groups)
3. Shanghai, Shanghai B, Suzhou, Ningbo, Wenzhou (Norman’s 1988 Wu group)
4. Nanchang, Anyi (Norman’s 1988 Gan group)
5. Changsha, Shuangfeng (Norman’s 1988 Xiang group)
6. Yingshan, Wuhan, Ningxia, Chengdu, Beijing, Taiyuan, Yuci (Norman’s 1988 Mandarin group)

The minimal lateral network based on the *Shùxíngtú* reference tree only differs from this network in that Guangzhou does not form a group with any other languages, while Meixian, Liancheng, Nanchang, and Anyi form a group. Given that List’s (2015) minimal lateral networks have an

almost identical structure to Ben Hamed & Wang's (2006) Neighbor Net, it is clear that the current majority tree deviates from these earlier results for the same dataset.

Given that the majority trees of the cov + top2 model and the cov + gam + top2 model are so different from earlier traditional and quantitative analyses, does this suggest that the majority trees pick up on affiliations between the languages that are non-tree-like? List (2015:40) presents the strongest lateral edges, connecting languages in different groupings. Let us compare the majority tree groupings to this table:

1. Wenzhou: Groups with the Min languages rather than the Wu languages. The reticulation drawing it towards the Min languages is clearly visible in Ben Hamed & Wang's (2006) Neighbor Net. List (2015:40) also notes a heavy lateral edge between Wenzhou and the Min, Hakka, and Yue languages.
2. Guangzhou: Groups with Meixian, one of two Hakka languages, rather than being by itself. Again, reticulation between these two languages is visible in Ben Hamed & Wang's (2006:54) Neighbor Net. Three of List's (2015:40) strongest lateral edges involve Guangzhou: those involving Wuhan, Suzhou, and Meixian.
3. Liancheng: Groups with Anyi and Nanchang, the two Gan languages, rather than with its Hakka sister Meixian. This connection is not evident from Ben Hamed & Wang (2006) or from List (2015), but may perhaps be an indirect consequence of the grouping of Guangzhou and Meixian.
4. Changsha and Shuangfeng (Xiang): Form a group with two of the Mandarin languages, Yingshan and Wuhan. These four languages are indeed close with considerable reticulation in Ben Hamed & Wang's (2006:54) Neighbor Net. The attraction between Changsha and the Mandarin languages is noted by List (2015:40) as well.
5. Ningbo: Groups with Ningxia and Chengdu, two of the Mandarin languages, rather than with the Wu languages. This reticulation is also present in Ben Hamed & Wang's (2006:54) Neighbor Net, and List (2015:40) notes a strong lateral edge between Ningbo and Mandarin, excluding Yingshan and Wuhan.

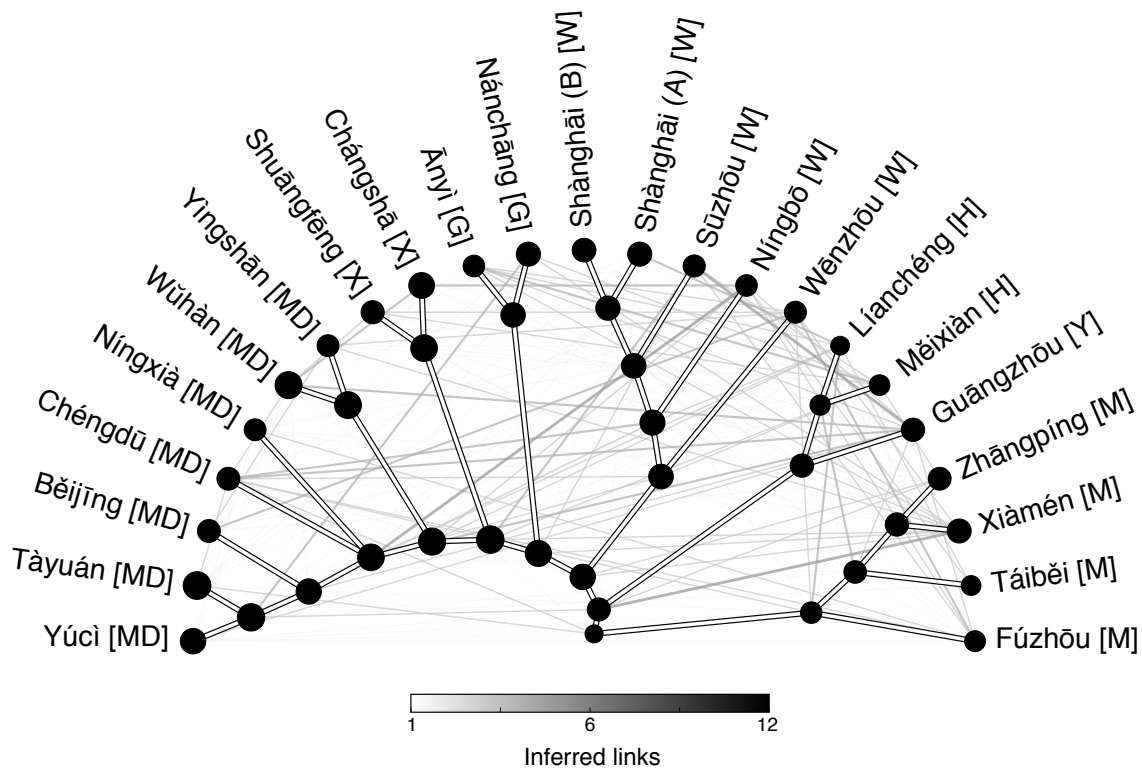


Figure 19. List's (2015) minimal lateral network using the *Arbre* reference tree. Reproduced with permission from the author. To allow for better legibility, the original figure has been changed so that the language labels are bigger and the legend has been moved

This discussion indicates that indeed, the groupings observed in the majority tree are caused by a signal in the data that is identified as reticulation in Ben Hamed & Wang's (2006: 54) Neighbor Net and as lateral edges in some of List's (2015) minimal lateral networks. What does the minority tree capture regarding the placement of these six languages?

1. Wenzhou: Groups with the Min languages rather than with the Wu languages in the majority tree. In the minority tree, it groups with the Wu languages Suzhou, Ningbo, Shanghai, and Shanghai B; this is the traditional genealogical grouping retrieved by Ben Hamed & Wang's (2006:54) Neighbor Net and List's (2015) minimal lateral networks.
2. Guangzhou: Groups with Meixian, one of two Hakka languages in the majority tree. In the minority tree, it is by itself, but still close to Meixian.
3. Liancheng, the other Hakka language: Groups with Anyi and Nanchang, the two Gan languages in the majority tree. In the minority tree, it is much closer to Meixian, but the two are not sisters as they are in Ben Hamed & Wang's (2006) Neighbor Net.
4. Changsha and Shuangfeng (Xiang): Form a group with two of the Mandarin languages, Yingshan and Wuhan, in the majority tree. In the minority tree, Changsha is located just outside of the Mandarin group, followed by Nanchang and then by Shuangfeng. This indicates that Changsha has a closer affinity to the Mandarin languages than Shuangfeng does.
5. Ningbo: Groups with Ningxia and Chengdu, two of the Mandarin languages, in the majority tree. In the minority tree, it groups with the Wu languages Suzhou, Wenzhou, Shanghai, and Shanghai B; this is the traditional genealogical grouping retrieved by Ben Hamed & Wang's (2006: 54) Neighbor Net and List's (2015) minimal lateral networks.

The minority tree picks up on the signals that are not present in the majority tree, especially with regard to the structure of the Wu language group, i.e. the placement of Wenzhou and Ningbo. However, where Ben Hamed & Wang (2006) and List (2015) mark the Wu group as a

valid vertical clade with Shanghai, Shanghai B, Suzhou, Ningbo, and Wenzhou, the multiple topologies method only recognizes this in the minority tree, not in the majority tree.

To further assess the contribution of different cognate sets to the weightings of the majority and minority trees, additional analyses using the “sitelh” command in *BayesPhylogenies* were conducted, as was done for Indo-European and Japonic. It is possible in this analysis to check this data against the meanings identified by List (2015:40) that heavily contribute to the lateral edges in his minimal lateral networks. For the cov + gam + top2 analysis, it was found that almost all cognate sets offer support both for the majority tree and for the minority tree across all iterations of the MCMC chain: 1460 sites (out of 1511) offer support for the majority tree, and 1482 sites offer support for the minority topology (similar to what was found for Indo-European). There are some sites that are only associated with one of the two trees: 51 sites are associated only with the majority topology, and 29 sites are associated only with the minority topology. The language groupings made by these 51 and 29 cognate sets, unfortunately, do not present any discernible pattern. Comparing these cognate sets to cognate sets associated only with the majority and minority trees for the cov + top2 model, only one site is found to exclusively support the minority topology, while 25 sites support only the majority topology. These sites are a subset of the sites identified as majority-supporting only for the cov + gam + top2 model. They do not overlap with List’s (2015:40) overview of meanings that contribute heavily to the lateral edges found in his minimal lateral networks. It is unfortunate that a clear explanation does not arise; however, it is interesting that the two models find such a narrowly-defined set of cognate sets that support only the minority tree.

5. Discussion

In this article, the behavior of the multiple topologies method implemented in *BayesPhylogenies* has been described for four lexical datasets, Austronesian (Gray, Drummond & Greenhill 2009), Indo-European (Bouckaert et al. 2012), Japonic (Lee & Hasegawa 2011), and Sinitic (Ben Hamed & Wang 2006). These datasets have all displayed rather different behavior; in all cases, it is impossible to point to the exact reasons for the non-tree-like signal that was found, given the large number of processes that are responsible for change in lexical datasets. The results so far indicate that there has been no evidence whatsoever for a second topology for the Austronesian dataset.

There has been support for alternative histories in the Indo-European dataset. The two topologies found for the highest ranking cov + gam + top2 model look quite similar: in both the majority and minority MCC trees (Figures 7 and 8), the main subfamilies of Indo-European are correctly identified, and there are no languages with differing subgroup affiliation between the trees, as was observed for Japonic and Chinese (with the exception of Oscan, Umbrian, Luvian, and Lycian, which is due to large amounts of missing data). Hence no clear evidence for reticulation driving support for a minority topology was found. The main driver of the well-supported minority tree set seems to be a large group of cognate sets with larger rates of change that have reflexes in Germanic, Romance, and Slavic languages. Greater branch lengths both leading to and within these subfamilies are evident from the minority MCC tree, but this is especially clear from the Neighbor Net conducted on only those cognate sets that have a support for the minority topology of 5 log-likelihood units or larger (Figure A.4). As a result of this split, the branches of the Germanic, Romance, and Slavic subfamilies in the majority tree seem rather stunted, and so this tree does a worse job than the minority tree does in identifying subfamily groupings such as Northwest Germanic and South Slavic.

Indo-European is a unique language family because of its well-attested older and ancient languages, of which 22 (including Cornish) are included in the current analyses (see Chang et al.

2015:205, 214–215 on special characteristics of ancient languages in phylogenetic analyses). Their placement within the minority and majority tree is not always correct, with the majority MCC tree especially showing a tendency to place non-contemporary languages close to each other. This trend is seen in the groupings of Avestan, Old Persian, and Vedic Sanskrit, as well as the grouping of Old High German and Gothic; similarly, Old Church Slavonic is pulled out of South Slavic. It may be the case that during cognate-coding, it is more straightforward to find cognates between ancient languages than it is to find cognates with modern reflexes due to the way etymological materials are organized. But this is not a problem in normal phylogenetic analyses, where older and ancient languages are constrained through calibration points. Chang et al. (2015: 215) conclude that the ancient languages Hittite, Vedic Sanskrit, Avestan, Ancient Greek, and Latin do not have more word forms per concept than modern languages do (their dataset is very similar to the current one). If anything, then, the emphasis on links between older or ancient languages does not result in them having significantly more synonyms for a given concept.

The Japonic dataset, likewise, was found to support more than one topology. It consists of data drawn from highly related languages, and as noted earlier, Lee & Hasegawa's (2011) maximum clade credibility tree is not well-supported for several internal nodes. Nor does it retrieve the clear division between east and west Japonic languages established in traditional Japanese linguistics (Shibatani 1990:196ff). The same was true for the current analysis, with the majority trees capturing some groupings not in Lee & Hasegawa's (2011) maximum clade credibility tree and the minority trees model capturing some language groupings that were not evident from the majority topology. Looking at the differences between Lee & Hasegawa's (2011) maximum clade credibility tree and the majority trees of the cov + gam + top2 and the cov + top2 models, it is helpful to consider the geographic location of the different languages given in Figure 20. Interesting differences between Lee & Hasegawa's (2011) analysis and the majority trees included the following:

1. The position of Shiga (number 25 in Figure 20), which is included in a clade with Nara (29), Kyoto (26), Osaka (28), and Wakayama (30) in the majority trees, similar to in Lee & Hasegawa's (2011:6) Neighbor Net. Shiga is geographically very close to these languages. Shiga's position in Lee & Hasegawa (2011) is not well supported, but close to Hyogo (32), Tokushima (38), Okayama (33), and Hiroshima (35). It is unclear which of these groupings captures vertical or horizontal signal. In the minority trees, Shiga is positioned close to Osaka too.
2. The position of Hiroshima (number 35 in Figure 20), which is included in the Hyogo-Tokushima-Kagawa-Ehime-Kochi-Okayama group in Lee & Hasegawa's (2011) analysis but is placed just outside of that group in the majority trees. It seems that this latter position is caused by Hiroshima being attracted to the Southern Kyushu languages (in blue in Figure 20) – this is also clear from Lee & Hasegawa's (2011: 6) Neighbor Net.
3. The group Chiba (number 15 in Figure20)-Gunma (11)-Saitama (12)-Kanagawa (14), which is found in the majority trees but not in Lee & Hasegawa's (2011) maximum clade credibility tree. These are all very close geographically (see Figure 20). In Lee & Hasegawa's (2011) analysis, Saitama (12) and Gunma (11) group with Nagano (16), Nigata (7), Hachijyo (21), and Shizuoka, but as in the grouping in the majority trees, this is not supported in more than 50% of the tree set. Saitama patterns with Tokushima (38) and Nagano (16) in the minority trees.
4. Nigata (number 7 in Figure 20)-Toyama (22)-Ishikawa (23)-Fukui (24)-Nagano (16), a grouping in the majority trees not found in Lee & Hasegawa (2011). It groups together the Western and Eastern languages that are spoken on the northern coast, so these languages are geographically close. In Lee & Hasegawa's (2011) consensus tree, Nigata (7) and Nagano (16) form a group together with Hachijyo (21), Shizuoka (18), Saitama

(12), and Gunma (11). In the majority tree, Hachijyo (21) and Shizuoka (18) are pulled outside of the Nigata-Toyama-Ishikawa-Fukui-Nagano group, and Saitama (12) and Gunma (11) group with Chiba (15) and Kanagawa (14). None of these groups are well-supported, either in the majority trees or in Lee & Hasegawa (2011), suggesting a large amount of uncertainty.

What can we speculate about the processes underlying the reticulation found in the Japonic dataset? These are all highly related language varieties that started to diverge only about 2,400 years ago, which makes a process like long branch attraction unlikely to have any effect. Japonic languages spread from Kyushu, the southernmost of Japan's large islands, and from there split up into what are called the mainland Japanese languages on Kyushu, Shikoku, Honshu, and Hokkaido; and the Ryukyuan languages on the smaller islands located to the south (Robbeets 2015:27ff). The absence of a clear east-west split and instead the grouping of languages that are in close proximity, such as the Nigata-Toyama-Ishikawa-Fukui-Nagano group in all analyses, suggests that borrowing or a slow splitting up of dialect chains may be the cause (Lee & Hasegawa 2011:6). Lee & Hasegawa (2011:4) note that the number of isoglosses separating languages is small and there is not much overlap; additionally, they state that mainland Japanese languages remained in contact through the road network. Further evidence for contact within the Japanese islands and diversification across islands is provided by Lee & Hasegawa (2014).

This is, of course, speculation, but it seems that there is some evidence for incomplete lineage sorting as well. A few patterns in the data suggest processes other than dialect chain break up and/or (continued) borrowing. Incomplete lineage sorting occurs when there are ancestral polymorphisms that are by chance inherited only partially by different species or languages, possibly in a scattered manner. A possible case could be the well-supported connection between Toyama and Okayama in the minority trees, as well as the connection within the set of languages that share a reflex for one of the cognate sets for 'tongue', illustrated by the red circles in Figure 20. The cognate sets for 'tongue' serve to group together languages spoken far apart in non-adjacent geographical areas, making diffusion or borrowing unlikely. An alternative explanation could be influence from the political and religious capital between 700 and 1868, which was located first in Nara (number 29 in Figure 20) and then in Kyoto (26) – both in Central Japan – although neither of these languages share this particular reflex for 'tongue'. While both the Kyoto and Tokyo languages, as languages of the capital, served as *lingua francas*, and although the latter served as the basis for the formation of the standard language (Shibatani 1990:185–187), neither seems involved in much reticulation for the current dataset. Lee & Hasegawa's (2011) Neighbor Net does not suggest this, and neither do the Delta scores and Q-Residuals calculated from the Neighbor Net analysis (as performed by the current author). Nevertheless, the influence of the old capitals as centers of innovation, pushing western characteristics outwards, should not be disregarded as a potential alternative explanation for these scattered patterns (this process is well attested in traditional Japanese linguistics; see Shibatani 1990:200ff).

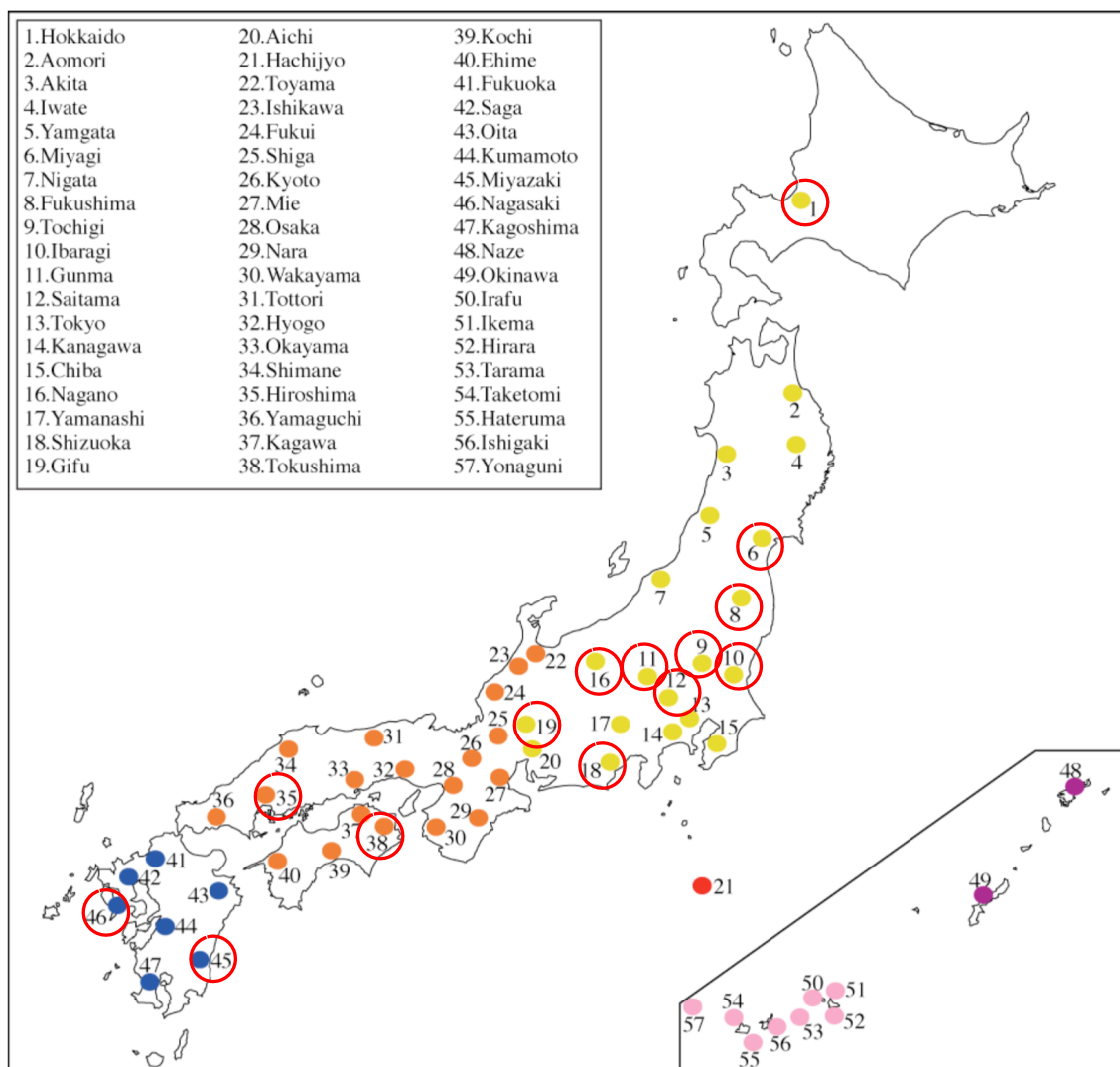


Figure 20. Map detailing the locations³ of the contemporary languages taken from Lee & Hasegawa (2011) (their Figure s1). From Lee, Sean, & Toshikazu Hasegawa. 2011. Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proceedings of the Royal Society B*, 278:1725.3662–3669. By permission of the Royal Society. Legend: yellow = eastern Japanese; orange = western Japanese; red = Hachijyo; blue = Kyushu; purple = northern Ryukyuan; pink = southern Ryukyuan.

The Sinitic dataset shows the most dramatic division in majority and minority tree weightings, with an almost equal split between sites. Analyses of which sites support which tree revealed that across iterations, almost all sites could support both the majority and the minority topology. It was already clear from earlier studies (Ben Hamed & Wang 2006; List 2015) that a large amount of uncertainty and reticulation is present in this dataset. Ben Hamed & Wang (2006) find that parsimony trees might find appropriate subgroups but have poor bootstrap support, while Neighbor Nets capture both this reticulation and the major subdivisions identified by the trees. This is due to two opposing forces in Sinitic language history: one of differentiation, in which people spreading throughout China established new communities and languages; and one of homogenization, in which wide-spread multilingualism and borrowing created linkages

³ The red circles were added by the current author and mark languages that have reflexes of one of two cognate sets for the meaning 'tongue'; see text.

between the languages (Ben Hamed & Wang 2006:53). Norman (1988:185) describes the first process as a “centrifugal force,” and the second as a “centripetal counterbalance.” The standard languages of the Chinese dynasties have influenced all Sinitic languages throughout most of their histories, but at the same time, they have had different impacts on different languages (Norman 1988: 185ff, Wang 1997: 57).

It is helpful to look at the geographic position of the languages, and so a map giving the position of Ben Hamed & Wang’s (2006) sampled languages is presented in Figure 21. Several of the points raised earlier regarding the majority trees and how they deviate from Ben Hamed & Wang (2006) and List (2015) might be relevant to issues of borrowing between closely situated languages or to major migrations (see Zhou 1991 on migrations). Wenzhou (number 15 in Figure 21), which is affiliated in the majority trees with the Min languages rather than with the Wu languages proposed by Ben Hamed & Wang (2006) and List (2015), is closest to the Min languages geographically, as it borders with Fuzhou (18). Changsha (8) and Shuangfeng (9) (Xiang) form a group with two of the Mandarin languages, Yingshan (7) and Wuhan (3), in the majority trees. These four languages are indeed situated quite close together. Clearly, not all divergences away from the topologies found in Ben Hamed & Wang (2006) and List (2015) can be explained by geographical proximity. A case in point is Ningbo (12) (Wu), whose closest affiliation in the majority trees is with Ningxia (4) and Chengdu (5) (Mandarin): these are far apart on the longitudinal axis. Zhou (1991) describes in detail three major migrations that have contributed to language formation. Most of these migrations took place from north to south and/or from west to east. Zhou (1991: 41) mentions one particular migration towards the coast of what is now Jiangsu and Zhejiang from the north and specifically mentions northern migrants settling in Ningbo.

The analyses of the four language families show a two-way split: on the one hand, we have Austronesian and Indo-European, which have no evidence for multiple topologies or evidence that does not point towards language contact; on the other hand, we have Japonic and Sinitic, which have closely related language varieties and where continuous opposing forces of language diversification and homogenization through language contact and standard language use seem to have created reticulate patterns. The analyses show that it is worthwhile to consider using methods that aim to capture this reticulation even for datasets where known borrowings have been removed, because (i) we are very unlikely to remove all borrowings, especially more ancient ones (see Chang et al. 2015: 205), and (ii) borrowing is not the only process through which reticulate patterns may arise.

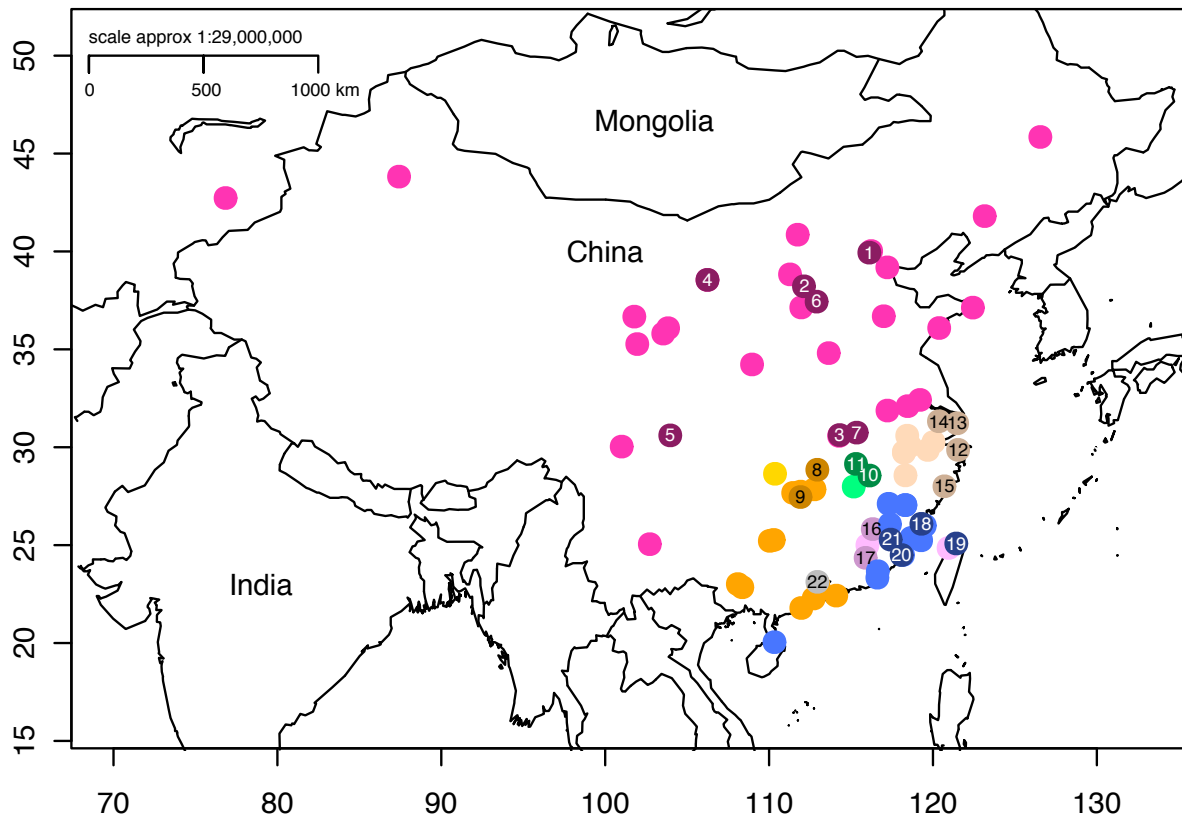


Figure 21. A map of the locations of the Sinitic languages, taken from Glottolog (Hammarström et al. 2017). Legend: **Mandarin (or Northern)** (light/dark pink): (1) Beijing, (2) Taiyuan, (3) Wuhan, (4) Ningxia, (5) Chengdu, (6) Yuci, (7) Yingshan. **Xiang** (light/dark orange): (8) Changsha, (9) Shuangfeng. **Gan** (light/dark green): (10) Anyi, (11) Nanchang. **Wu (or Wu-Hui)** (light/dark beige): (12) Ningbo, (13) Shanghai, (14) Suzhou, (15) Wenzhou. **Hakka** (light/dark lilac): (16) Liancheng, (17) Meixian. **Min** (light/dark blue): (18) Fuzhou, (19) Taiwan, (20) Xiamen, (21) Zhangping. **Yue (or Cantonese)** (gray): (22) Guangzhou. **Waxianghua** (gold).

Ben Hamed & Wang (2006: 55) make an interesting and valid point regarding the impact of data coding. Given that their coding (and this applies to the coding of lexical data in almost all phylogenetic endeavors) does not distinguish between change in a suffix and a change in the root, does not contain transitional states such as variant roots, and carries no information about the likelihood of the direction of change, it could be the case that the tree-like signal is “hidden” by the “neutral” coding of lexical data. This point has also been made by List (2016), who studies the effects of coding partial cognacy in different ways and pushes for the adoption of multistate models over the binary gain-lose models most commonly used now (see also List 2015:35 for an example of this problem in Ben Hamed & Wang’s 2006 dataset). This is a more general problem with the application of phylogenetic methods in linguistics. Furthermore, this fact would not have an equal effect on all four datasets considered in this study, due to the different morphological make-ups of languages in different language families.

Another issue regarding data collection and coding can be detected by dividing the number of cognate sets by the number of meanings and subsequently by the number of languages in the sample. For Austronesian, Sinitic, and Indo-European, the resulting number of cognate sets per

⁴ Light and dark hues of the same color indicate subfamily and numbered points in darker color indicate varieties present in Ben Hamed & Wang (2006). In order to increase legibility, the distances between the points representing Taiyuan and Yuci, Anyi and Nanchang, and Liancheng and Zhangping were increased slightly.

language and meaning is similar: 0.41, 0.31, and 0.27, respectively. For Japonic, this number is radically lower: 0.05 cognate sets per language per meaning. This may be one of the reasons for Lee & Hasegawa's (2011) poorly-supported clades: it might be the case that not enough variation is present in the dataset. In Indo-European, the inclusion of ancient languages may lead to a proliferation of cognate sets, as it may be unclear which of several attested near-synonyms is the "Swadesh term" (this term refers to the single word, mapping onto a given concept, that is most common, non-specialized and non-bound; see Swadesh 1952: 457). The Bouckaert et al. (2012) dataset may suffer from cognate over-population: although this has not been quantified, the dataset may include non-Swadesh terms for some concepts. Looking at Dutch, it includes *waar*, *recht*, *rechtmatig* and *echt* for 'true', and *breed*, *wijd*, and *ruim* for 'wide'. In both cases, only one term is the true Swadesh term, respectively *waar* 'true' and *breed* 'wide'. Including three or four synonym words for a concept, each with their own cognate history, may potentially introduce a source of reticulate signal as well as inflate the time depth of certain groups (see Chang et al. 2015: 211 on "overloaded" concept slots). Note that this is just speculation: it is unclear whether this played a significant role in the Bouckaert et al. (2012) analysis; note also that the number of cognate sets per language and meaning (0.27) is lower than that of the Austronesian (0.41) and Sinitic (0.31) datasets. Future advances in our knowledge of the impact of coding decisions on synonyms, in the partial cognate coding proposed by List (2016), and in other models of language evolution, including phoneme level analysis (Hruschka et al. 2015), may possibly help efforts to find phylogenetic signal in the future. The current article has shown that the multiple topologies method has recovered, at least in part, the reticulation in lexical datasets that we knew about on the basis of distance-based phylogenetic inference, mostly Neighbor Net analyses, as reported in the four original articles and earlier follow-up studies. It also recovered more than one signal in the Indo-European analysis, which was speculated to be mostly due to differences in rate of lexical change. The Japonic and Sinitic results are not surprising: (i) the reticulate nature of the Sinitic dataset has been discussed previously, and (ii) the original analysis of the Japonic dataset results in a consensus tree that is not always well supported, suggesting the presence of reticulation. After describing the results, I speculated on the potential sources of reticulation, describing differential break-up of dialect chains and borrowing, both shortly after language/dialect break-up and long after in Japonic and Sinitic. There may be evidence for incomplete lineage sorting in Japonic too. The Indo-European results show that even when borrowing is not available as an explanation, the multiple topologies method can pick up on other processes that generate multiple signals. However, as the current study is not a strict test of the multiple topologies method, it is not possible to generalize in any way the current results and provide guidelines on how and when to use the multiple topologies method, nor on the interpretation of the results if multiple topologies are found.

How does the multiple topologies method compare with the method proposed by Nakhleh, Ringe & Warnow (2005) and the minimal lateral network method (for instance, List 2015)? The big difference is that the latter informs the user on what is vertical and what is horizontal evolution (as reference phylogenies are used), while the multiple topologies method cannot do this. This makes the multiple topologies method less prone to biases introduced by the reference phylogeny, but some users might prefer to "know" which aspects of the two evolutionary histories that are recovered constitute genealogical descent, and which can be attributed to borrowing and other horizontal processes. For the Sinitic dataset, the minority trees were found to pick up the genealogical grouping of the Wu languages, while the majority trees group Wenzhou and Ningbo together with other languages from other subgroups. The majority trees for Sinitic thus seemingly capture an amalgamate of vertical and horizontal signal, while only the minority trees capture the true vertical history of the Wu group. Of course, this is not a surprising result either. After all, the multiple topologies method does not know anything about the different processes of language evolution that are distinguished here; it only knows that some

sites support one tree topology and other sites support another. In this respect, it is similar to other tools that are used to assess non-tree-like signal, such as Neighbor Nets and minimal lateral networks. However, the multiple topologies method outperforms these in two ways. First of all, it does not require a reference topology, thus taking one source of possible bias or error away from the analysis. Second, it is a character-based rather than a distance-based method, thus providing the user with the different paths along which different sites have changed, rather than summarizing average distances between language pairs, as Neighbor Net does.

6. Conclusion

In this article, I have described the first results of the use of the multiple topologies method (Pagel & Meade 2006a) implemented in *BayesPhylogenies* (Pagel & Meade 2004) with linguistic datasets. Four different language families were involved: Austronesian (data from Gray, Drummond & Greenhill 2009), Sinitic (Ben Hamed & Wang 2006), Indo-European (Bouckaert et al. 2012), and Japonic (Lee & Hasegawa 2011). Evidence for non-tree-like signal was found in all except Austronesian. Speculations on the origins of the non-tree-like signal were made in terms of a number of processes, including dialect chain break-up, borrowing (both shortly after language splits and later on), incomplete lineage sorting, and characteristics of lexical datasets. The Indo-European results suggest that rate of change differences between cognate sets can also be picked up as differentiating signals. The multiple topologies method is a useful tool for studying the dynamics of language evolution and may be used to help identify non-tree-like signal. However, as this article constitutes only an exploratory experiment, more formal tests of its behavior and applicability to linguistic datasets are needed.

Acknowledgements

First and foremost, I would like to thank Mark Pagel and Andrew Meade for their assistance on this project, especially Andrew, who helped to do analysis far beyond the call of duty. I would also like to thank Michael Dunn for preparing a new version of his Indo-European dataset to use in this paper, Sean Lee for making his Japonic dataset available, and Mattis List for help with the Sinitic dataset. Comments from three anonymous reviewers helped to make this a better paper, for which thanks are given. Thanks to Marcel Sommer for technical support. I thank the audience of ICHL22, the organizers of the workshop “Non-cladistic approaches to language genealogy,” Siva Kalyan and Alexandre François, and additional guest editor Harald Hammarström for their commitment. This research was supported by ERC grant ‘The evolution of human languages’ (reference number 268744).

References

- Algeo, John & Thomas Pyles. 2005. *The Origins and Development of the English Language (Sixth Edition)*. Boston: Wadsworth.
- Atkinson, Quentin D. & Russell D. Gray. 2005. Curious Parallels and Curious Connections: Phylogenetic Thinking in Biology and Historical Linguistics. *Systematic Biology* 54:4.513–526. <https://doi.org/10.1080/10635150590950317>
- Beekes, Robert S. P. 2011. *Comparative Indo-European Linguistics: An Introduction (Second Edition)*. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.172>
- Ben Hamed, Mahé & Feng Wang. 2006. Stuck in the Forest: Trees, Networks, and Chinese Dialects. *Diachronica* 23:1.29–60. <https://doi.org/10.1075/dia.23.1.04ham>

- Blench, Roger. 2015. “New Mathematical Methods” in Linguistics Constitute the Greatest Intellectual Fraud in the Discipline Since Chomsky. Talk presented at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.
- 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>
- Bowern, Claire & Quentin D. 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 D. 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>
- Bowern, Claire & Bethwyn Evans, eds. 2014. *The Routledge Handbook of Historical Linguistics*. London: Routledge.
- 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>
- 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>
- Chousou-Polydouri, Natalia, Joshua Birchall, Sérgio Meira, Zachary O’Hagan & Lev Michael. 2016. A Test of Coding Procedures for Lexical Data with Tupí-Guaraní and Chapacuran Languages. *Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics* ed. by Christian Bentz, Gerhard Jäger & Igor Yanovich. Tübingen: University of Tübingen. Available at <https://publikationen.uni-tuebingen.de/xmlui/handle/10900/68558>
- Collard, Mark, Stephen J. Shennan & Jamshid J. Tehrani. 2006. Branching, Blending, and the Evolution of Cultural Similarities and Differences Among Human Populations. *Evolution and Human Behavior* 27:3.169–184. <https://doi.org/10.1016/j.evolhumbehav.2005.07.003>
- Croft, William. 2008. Evolutionary Linguistics. *Annual Review of Anthropology* 37.219–234. <https://doi.org/10.1146/annurev.anthro.37.081407.085156>
- Drummond, Alexei J. & Remco R. Bouckaert. 2015. *Bayesian Evolutionary Analysis with BEAST*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139095112>
- Dunn, Michael. 2014. Language Phylogenies. In Claire Bowern & Bethwyn Evans, eds., 190–211. London: Routledge.
- Dunn, Michael. 2014. *Indo-European Lexical Cognate Data*. Zenodo. Available at <https://doi.org/10.5281/zenodo.1284236>
- 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* 473.79–82. <https://doi.org/10.1038/nature09923>

- Dyen, Isidore, Joseph B. Kruskal & Paul Black. 1992. An Indo-European Classification: A Lexicostatistical Experiment. *Transactions of the American Philosophical Society* 82.1–132. <https://doi.org/10.2307/1006517>
- Edwards, Anthony W. F. & Luigi Luca Cavalli-Sforza. 1964. Reconstruction of Evolutionary Trees. *Phenetic and Phylogenetic Classification* ed. by V. H. Heywood & J. McNeill, 67–76. London: Systematics Association.
- Eska, Joseph F. & Don Ringe. 2004. Recent Work in Computational Linguistic Phylogeny. *Language* 80:3.569–582. <https://doi.org/10.1353/lan.2004.0123>
- Forster, Peter & Colin Renfrew, eds. 2006. *Phylogenetic Methods and the Prehistory of Languages*. Cambridge: McDonald Institute Monographs.
- François, Alexandre. 2014. Trees, Waves, and Linkages: Models of Language Diversification. In Claire Bowerman & Bethwyn Evans, eds., 161–189.
- Kalyan, Siva & Alexandre François. 2018. Freeing the Comparative Method from the Tree Model: A Framework for Historical Glottometry. In Ritsuko Kikusawa & Lawrence Reid (eds.), *Let's talk about trees: Genetic Relationships of Languages and Their Phylogenetic Representation* (Senri Ethnological Studies, 98). Ōsaka: National Museum of Ethnology. 59–89.
- Galucio, Ana Vilacy, Sérgio Meira, Joshua Birchall, Denny Moore, Nilson Gabas Júnior, Sebastian Drude, Luciana Storto, Gessiane Picanço & Carmen Reis Rodrigues. 2015. Genealogical Relations and Lexical Distances Within the Tupian Linguistic Family. *Boletim do Museu Paraense Emílio Goeldi, Ciências Humanas* 10:2.229–274. <https://doi.org/10.1590/1981-81222015000200004>
- Geisler, Hans & Johann-Mattis List. 2013. Do Languages Grow on Trees? The Tree Metaphor in the History of Linguistics. *Classification and Evolution in Biology, Linguistics and the History of Science: Concepts – Methods – Visualization* ed. by Heiner Fangerau, Hans Geisler, Thorsten Halling & William Martin, 111–124. Stuttgart: Franz Steiner Verlag.
- Gray, Russell D., David Bryant & Simon J. Greenhill. 2010. On the Shape and Fabric of Human History. *Philosophical Transactions of the Royal Society B* 365:1559.3923–3933. <https://doi.org/10.1098/rstb.2010.0162>
- Gray, Russell D., Simon J. Greenhill & Robert M. Ross. 2007. The Pleasures and Perils of Darwinizing Culture (With Phylogenies). *Biological Theory* 2:4.360–375. <https://doi.org/10.1162/biot.2007.2.4.360>
- Gray, Russell D. & Fiona M. Jordan. 2000. Language Trees Support the Express-Train Sequence of Austronesian Expansion. *Nature* 405.1052–1055. <https://doi.org/10.1038/35016575>
- 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>
- Greenhill, Simon J., Thomas E. Currie & Russell D. Gray. 2009. Does Horizontal Transmission Invalidate Cultural Phylogenies? *Proceedings of the Royal Society B* 276.2299–2306. <https://doi.org/10.1098/rspb.2008.1944>
- 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>
- Hammarström, Harald, Sebastian Bank, Robert Forkel & Martin Haspelmath. 2017. *Glottolog 3.1*. Jena: Max Planck Institute for the Science of Human History.

- Heggarty, Paul. 2006. Interdisciplinary Indiscipline? Can Phylogenetic Methods Meaningfully be Applied to Language Data – And to Dating Language? In Peter Forster & Colin Renfrew, eds., 183–194.
- Heggarty, Paul, Warren Maguire & April McMahon. 2010. Splits or Waves? Trees or Webs? How Divergence Measures and Network Analysis Can Unravel Language Histories. *Philosophical Transactions of the Royal Society B* 365:1559.3829–3843. <https://doi.org/10.1098/rstb.2010.0099>
- Holm, Hans J. 2007. The New Arboretum of Indo-European “Trees”: Can New Algorithms Reveal the Phylogeny and Even Prehistory of Indo-European? *Journal of Quantitative Linguistics* 14:2–3.167–214. <https://doi.org/10.1080/09296170701378916>
- Hruschka, Daniel J., Simon Branford, Eric D. Smith, Jon Wilkins, Andrew Meade, Mark Pagel & Tanmoy Bhattacharya. 2015. Detecting Regular Sound Changes in Linguistics as Events of Concerted Evolution. *Current Biology* 25:1.1–9. <https://doi.org/10.1016/j.cub.2014.10.064>
- Huelsenbeck, John P., Fredrik Ronquist, Rasmus Nielsen & Jonathan P. Bollback. 2001. Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. *Science* 294:5550.2310–2314. <https://doi.org/10.1126/science.1065889>
- Huson, Daniel H. & David Bryant. 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* 23:2.254–267. <https://doi.org/10.1093/molbev/msj030>
- Huson, Daniel H. & Celine Scornavacca. 2010. A Survey of Combinatorial Methods for Phylogenetic Networks. *Genome Biology and Evolution* 3.23–35. <https://doi.org/10.1093/gbe/evq077>
- Huson, Daniel H. & Celine Scornavacca. 2012. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology* 61:6.1061–1067. <https://doi.org/10.1093/sysbio/sys062>
- Jacques, Guillaume & Johann-Mattis List. This issue. Save the Trees: Why We Need Tree Models in Linguistic Reconstruction. *Journal of Historical Linguistics* 9:1. 128–167.
- Jaeger, Gerhard & Søren Wichmann. 2016. Inferring the World Tree of Languages From Word Lists. *The Evolution of Language: Proceedings of the 11th International Conference on the Evolution of Language (EvoLang XI)* ed. by Séan G. Roberts, Christine Cuskley, Luke McCrohon, Lluís Barceló-Coblijn, Olga Feher & Tessa Verhoef. Available at <http://evolang.org/neworleans/papers/147.html>
- Kalyan, Siva & Alexandre François. 2018. Freeing the Comparative Method from the tree model: A framework for Historical Glottometry. In Ritsuko Kikusawa & Lawrence Reid (eds), *Let's talk about trees: Genetic Relationships of Languages and Their Phylogenetic Representation (Senri Ethnological Studies, 98)*. Ōsaka: National Museum of Ethnology. 59–89.
- Kelly, Luke J. & Geoff K. Nicholls. 2017. Lateral Transfer in Stochastic Dollo Models. *The Annals of Applied Statistics* 11:2.1146–1168. <https://doi.org/10.1214/17-AOAS1040>
- Koonin, Eugene V., Kira S. Makarova & L. Aravind. 2001. Horizontal Gene Transfer in Prokaryotes: Quantification and Classification. *Annual Review of Microbiology* 55:1.709–742. <https://doi.org/10.1146/annurev.micro.55.1.709>
- Lee, Sean. 2018. *The Best Fitting Tree Model for Japonic Languages*. Zenodo. Available at <https://doi.org/10.5281/zenodo.1220896>
- Lee, Sean & Toshikazu Hasegawa. 2011. Bayesian Phylogenetic Analysis Supports an Agricultural Origin of Japonic Languages. *Proceedings of the Royal Society B*, 278.3662–3669. <https://doi.org/10.1098/rspb.2011.0518>

- Lee, Sean & Toshikazu Hasegawa. 2014. Oceanic Barriers Promote Language Diversification in the Japanese Islands. *Journal of Evolutionary Biology* 27:9.1905–1912. <https://doi.org/10.1111/jeb.12442>
- Levinson, Stephen C. & Russell D. Gray. 2012. Tools from Evolutionary Biology Shed New Light on the Diversification of Languages. *Trends in Cognitive Sciences* 16:3.167–173. <https://doi.org/10.1016/j.tics.2012.01.007>
- List, Johann-Mattis. 2015. Network Perspectives on Chinese Dialect History: Chances and Challenges. *Bulletin of Chinese Linguistics* 8:1.27–47. <https://doi.org/10.1163/2405478X-00801002>
- List, Johann-Mattis. 2016. Beyond Cognacy: Historical Relations Between Words and Their Implication for Phylogenetic Reconstruction. *Journal of Language Evolution* 1:2.119–136. <https://doi.org/10.1093/jole/lzw006>
- List, Johann-Mattis. 2017. *Chinese Dialect Database*. Jena: Max Planck Institute for the Science of Human History. Available at <https://github.com/digling/cddb>
- List, Johann-Mattis, Shijulal Nelson-Sathi, Hans Geisler & William Martin. 2013. 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, Shijulal Nelson-Sathi, William Martin & Hans Geisler. 2014. Using Phylogenetic Networks to Model Chinese Dialect History. *Language Dynamics and Change*, 4:2.222–252. <https://doi.org/10.1163/22105832-00402008>
- 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* 11.360–375. <https://doi.org/10.1186/s13062-016-0145-2>
- Macklin-Cordes, Jayden, & Erich Round. 2015. High-Definition Phonotactics Reflect Linguistic Pasts. *Proceedings of the 6th Conference on Quantitative Investigations in Theoretical Linguistics* ed. by Johannes Wahle, Marisa Kollner, Harald Baayen, Gerhard Jäger & Tineke Baayen-Oudshoorn, 1–5. Tübingen: University of Tübingen.
- Maddison, Wayne P. 1997. Gene Trees in Species Trees. *Systematic Biology* 46:3.523–536. <https://doi.org/10.1093/sysbio/46.3.523>
- Matthews, Luke J., Jamie J. Tehrani, Fiona M. Jordan, Mark Collard & Charles L. Nunn. 2011. Testing for Divergent Transmission Histories Among Cultural Characters: A Study Using Bayesian Phylogenetic Methods and Iranian Tribal Textile Data. *PLOS ONE* 6.e14810. <https://doi.org/10.1371/journal.pone.0014810>
- Meira, Sérgio, Joshua Birchall & Natalia Chousou-Polydouri. 2015. A Character-Based Internal Classification of the Cariban Language Family. Paper presented at the 48th Annual Meeting of the Societas Linguistica Europaea, Leiden, The Netherlands, September 2–5, 2015.
- Morrison, David A. 2016. Genealogies: Pedigrees and Phylogenies Are Reticulating Networks Not Just Divergent Trees. *Evolutionary Biology* 43:4.456–473. <https://doi.org/10.1007/s11692-016-9376-5>
- Nakhleh, Luay, Don Ringe & Tandy Warnow. 2005. Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages. *Language* 81:2.382–420. <https://doi.org/10.1353/lan.2005.0078>
- Nelson-Sathi, Shijulal, Johann-Mattis List, Hans Geisler, Heiner Fangerau, Russell D. Gray, William Martin & Tal Dagan. 2011. Networks Uncover Hidden Lexical Borrowing in

- Indo-European Language Evolution. *Proceedings of the Royal Society B* 278:1713.1794–1803. <https://doi.org/10.1098/rspb.2010.1917>
- Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.
- Nunn, Charles L., Christian Arnold, Luke Matthews & Monique Borgerhoff Mulder. 2010. Simulating Trait Evolution for Cross-Cultural Comparison. *Philosophical Transactions of the Royal Society B* 365:1559.3807–3819. <https://doi.org/10.1098/rstb.2010.0009>
- Nunn, Charles L., Monique Borgerhoff Mulder & Sasha Langley. 2006. Comparative Methods for Studying Cultural Trait Evolution: A Simulation Study. *Cross-Cultural Research* 40:2.177–209. <https://doi.org/10.1177/1069397105283401>
- Pagel, Mark. 2009. Human Language as a Culturally Transmitted Replicator. *Nature Reviews Genetics* 10.405–415. <https://doi.org/10.1038/nrg2560>
- Pagel, Mark, Quentin D. Atkinson & Andrew Meade. 2007. Frequency of Word-Use Predicts Rates of Lexical Evolution Throughout Indo-European History. *Nature* 449.717–721. <https://doi.org/10.1038/nature06176>
- Pagel, Mark & Andrew Meade. 2004. A Phylogenetic Mixture Model for Detecting Pattern-Heterogeneity in Gene Sequence or Character-State Data. *Systematic Biology* 53:4.571–581. <https://doi.org/10.1080/10635150490468675>
- Pagel, Mark & Andrew Meade. 2006a. Detecting Conflicting Phylogenetic Signal: A Mixture Model Approach with Multiple Tree Topologies. Paper presented at the New Zealand Phylogenetics Meeting, Kaikoura, New Zealand, February 12–17, 2006. Available at <https://zenodo.org/record/1117335#.WjeEzSOZPFR>
- Pagel, Mark & Andrew Meade. 2006b. Estimating Rates of Lexical Replacement on Phylogenetic Trees of Languages. In Peter Forster & Colin Renfrew, eds., 173–182.
- Pagel, Mark & Andrew Meade. N.d. Draft Manual for *BayesPhylogenies*. Available at <http://www.evolution.reading.ac.uk/BayesPhy.html>
- Pereltsvaig, Asya & Martin W. Lewis. 2015. *The Indo-European Controversy: Facts and Fallacies in Historical Linguistics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781107294332>
- Robbeets, Martine. 2015. *Diachrony of Verb Morphology: Japanese and the Transeurasian Languages*. Berlin: Walter de Gruyter. <https://doi.org/10.1515/9783110399943>
- Rokas, Antonis & Sean B. Carroll. 2006. Bushes in the Tree of Life. *PLOS Biology* 4:11.e352. <https://doi.org/10.1371/journal.pbio.0040352>
- Ronquist, Fredrik, Paul van der Mark & John P. Huelsenbeck. 2009. Bayesian Phylogenetic Analysis Using MrBayes. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing* ed. by Marco Salemi, Philippe Lemey & Anne-Mieke Vandamme, 210–266. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511819049.009>
- Shibatani, Masayoshi. 1990. *The Languages of Japan*. Cambridge: Cambridge University Press.
- Swadesh, Morris. 1952. Lexico-statistic Dating of Prehistoric Ethnic Contacts: With Special Reference for North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96:4.452–463.
- Swadesh, Morris. 1955. Towards Greater Accuracy in Lexicostatistic Dating. *International Journal of American Linguistics* 21:2.121–137. <https://doi.org/10.1086/464321>
- Tadmor, Uri. 2009. Loanwords in the World's Languages: Findings and Results. *Loanwords in the World's Languages: A Comparative Handbook* ed. by Martin Haspelmath & Uri Tadmor, 55–75. Berlin: Walter de Gruyter. <https://doi.org/10.1515/9783110218442.55>

- Tehrani, Jamshid & Mark Collard. 2002. Investigating Cultural Evolution Through Biological Phylogenetic Analyses of Turkmen Textiles. *Journal of Anthropological Archaeology* 21:4.443–463. [https://doi.org/10.1016/S0278-4165\(02\)00002-8](https://doi.org/10.1016/S0278-4165(02)00002-8)
- Verkerk, Annemarie. 2014. The Evolutionary Dynamics of Motion Event Encoding. Radboud University Nijmegen PhD dissertation.
- Wang, Feng. 2004. BCD: Basic Words of Chinese Dialects. Available at <https://github.com/digling/cddb/tree/master/datasets/Wang2004a>
- Wang, William S.-Y. 1997. Languages or Dialects? *The CUHK Journal of Humanities* 1.54–62.
- Wang, William S.-Y. & James W. Minett. 2005. Vertical and Horizontal Transmission in Language Evolution. *Transactions of the Philological Society* 103:2.121–146. <https://doi.org/10.1111/j.1467-968X.2005.00147.x>
- Wichmann, Søren, Robert S. Walker, Taraka Rama & Eric W. Holman. 2011. Correlates of Reticulation in Linguistic Phylogenies. *Language Dynamics and Change* 1:2.205–240. <https://doi.org/10.1163/221058212X648072>
- Widmer, Manuel, Sandra Auderset, Johanna Nichols, Paul Widmer & Balthasar Bickel. 2017. NP Recursion Over Time: Evidence from Indo-European. *Language* 93:4.799–826. <https://doi.org/10.1353/lan.2017.0058>
- Willems, Matthieu, Etienne Lord, Louise Laforest, Gilbert Labelle, François-Joseph Lapointe, Anna Maria Di Sciullo & Vladimir Makarenkov. 2016. Using Hybridization Networks to Retrace the Evolution of Indo-European Languages. *BMC Evolutionary Biology* 16:1.180. <https://doi.org/10.1186/s12862-016-0745-6>
- Xie, Wangang, Paul O. Lewis, Yu Fan, Lynn Kuo & Ming-Hui Chen. 2010. Improving Marginal Likelihood Estimation for Bayesian Phylogenetic Model Selection. *Systematic Biology* 60:2.150–160. <https://doi.org/10.1093/sysbio/syq085>
- Zhou, Kevin & Claire Bower. 2015. Quantifying Uncertainty in the Phylogenetics of Australian Numeral Systems. *Proceedings of the Royal Society B* 282:1815.20151278. <https://doi.org/10.1098/rspb.2015.1278>
- Zhou, Zhenhe. 1991. Migrations in Chinese History and Their Legacy on Chinese Dialects. *Journal of Chinese Linguistics Monograph Series* 3.29–49.

Appendix

Table 5. Model comparison for Bouckaert et al.'s (2012) updated lexical dataset on 116 Indo-European languages, adding the additional analysis including site log-likelihoods (in bold)

model	marginal l-lh	marginal l-lh difference	weight majority tree	weight minority tree
cov + gam + top2	-51167		0.60	0.40
cov + gam + top2 sitelh	-51179	12	0.55	0.45
cov + top2	-51182	3	0.65	0.35
gam + top2	-51375	193	0.77	0.23
cov + gam	-51450	75	-	-
cov	-51524	74	-	-
gam	-51791	267	-	-
top2	-51832	41	0.64	0.36
standard	-52812	980	-	-

a. marginal l-lh = highest marginal log-likelihood out of 5 runs

b. marginal l-lh difference = difference in log-likelihood units with better performing model in above row

c. weight majority tree & weight minority tree = for those models that estimate two topologies, the mean weight associated with the best supported and least supported tree, respectively

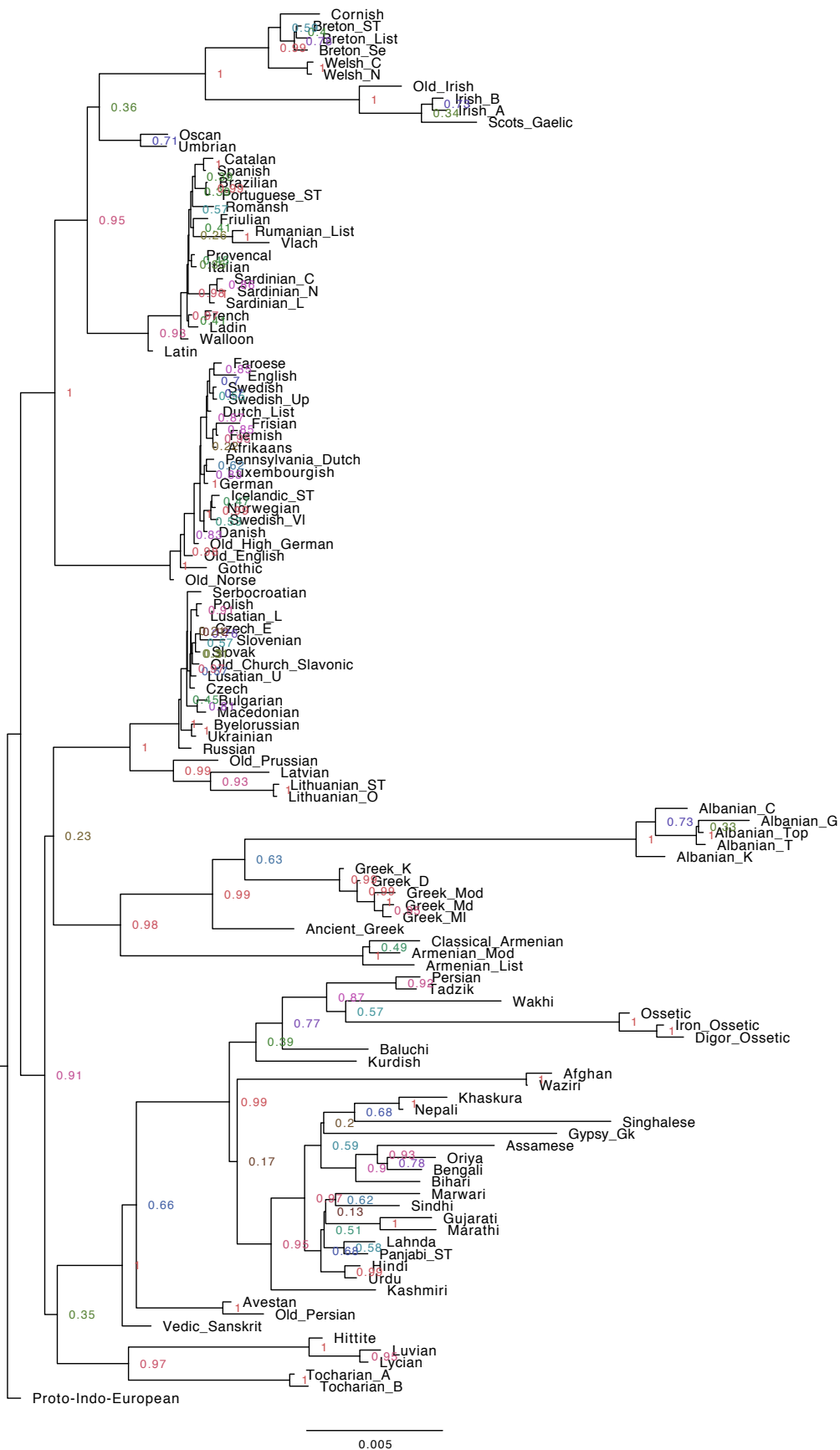


Figure 22. The majority maximum clade credibility tree for the cov + gam + top2 model with the highest marginal log-likelihood for the additional site log-likelihood analysis (Bouckaert et al. 2012)

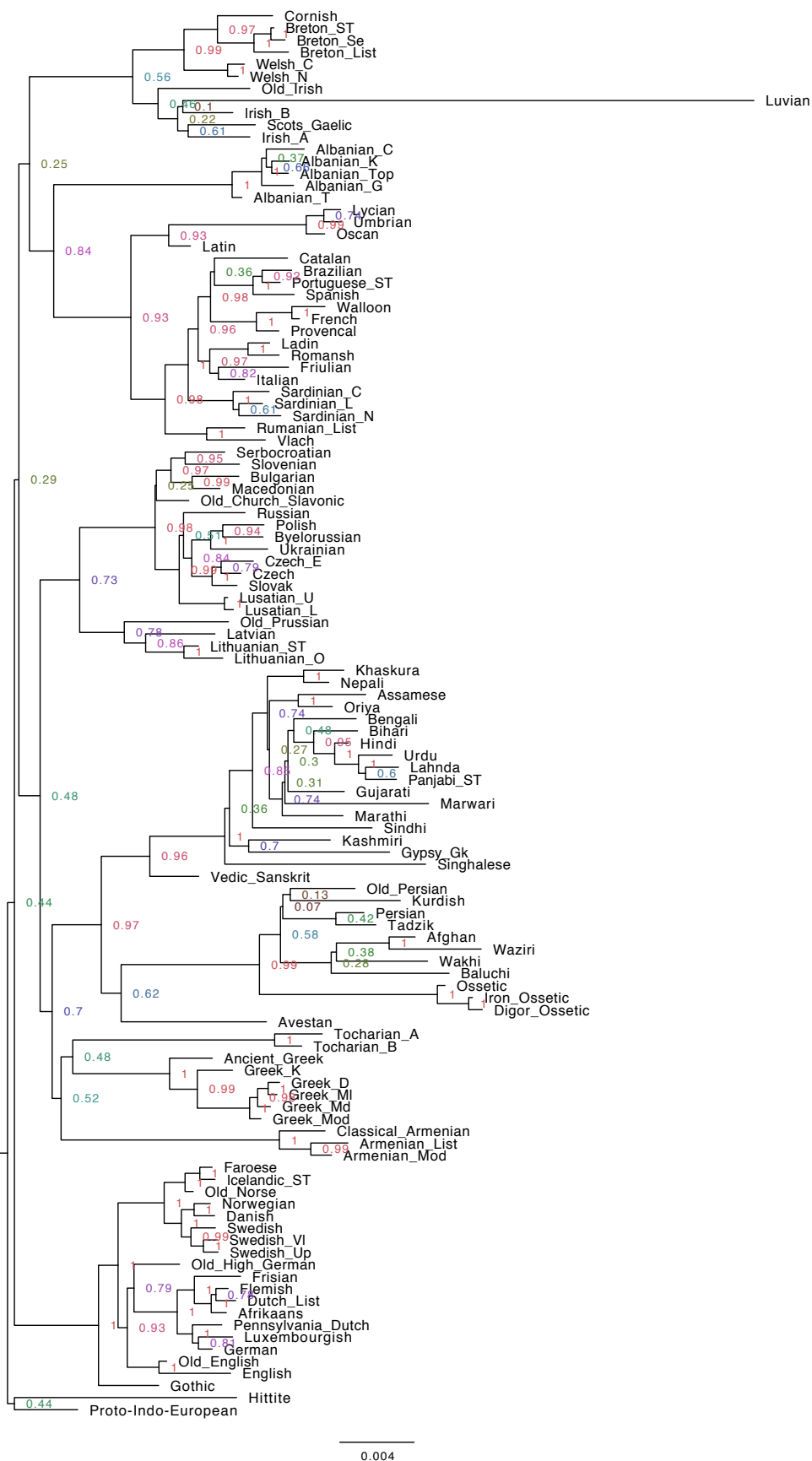


Figure 23. The minority maximum clade credibility tree for the cov + gam + top2 model with the highest marginal log-likelihood for the additional site log-likelihood analysis (Bouckaert et al. 2012)

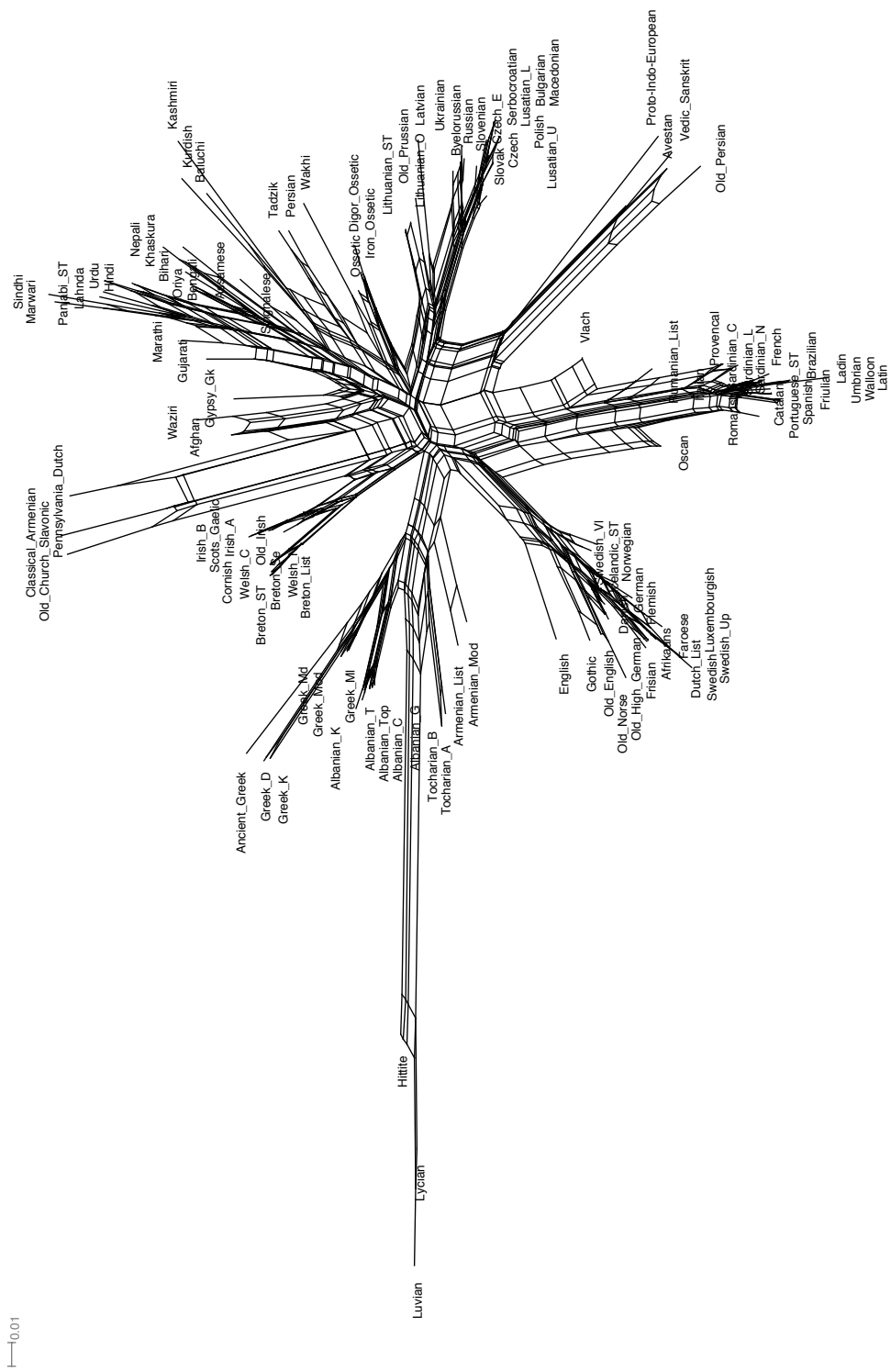


Figure 24. Neighbor Net including only the 97 cognate sets that show a supporting log-likelihood difference of 5 or more for the majority tree in analysis of Indo-European (Bouckaert et al. 2012)

Address for correspondence

Annemarie Verkerk
Max Planck Institute for the Science of Human History
Kahlaische Strasse 10
07745 Jena
Germany

verkerk@shh.mpg.de