Three tree priors and five datasets: A study of Indo-European phylogenetics

Taraka Rama
Department of Informatics
University of Oslo, Norway
tarakark@ifi.uio.no

Abstract

The age of the root of the Indo-European language family has received much attention since the application of Bayesian phylogenetic methods by Gray and Atkinson (2003). With the application of new models, the root age of the Indo-European family has tended to decrease from an age that supported the Anatolian origin hypothesis to an age that supports the Steppe origin hypothesis (Chang et al., 2015). However, none of the published work in Indo-European phylogenetics has studied the effect of tree priors on phylogenetic analyses of the Indo-European family. In this paper, I intend to fill this gap by exploring the effect of tree priors on different aspects of the Indo-European family's phylogenetic inference. I apply three tree priors – Uniform, Fossilized Birth-Death (FBD), and Coalescent – to five publicly available datasets of the Indo-European language family. I evaluate the posterior distribution of the trees from the Bayesian analysis using Bayes Factor, and find that there is support for the Steppe origin hypothesis in the case of two tree priors. I report the median and 95% highest posterior density (HPD) interval of the root ages for all three tree priors. A model comparison suggests that either the Uniform prior or the FBD prior is more suitable than the Coalescent prior to the datasets belonging to the Indo-European language family.

1 Introduction

The Indo-European language family is widely spoken and consists of languages belonging to subgroups such as Albanian, Armenian, Balto-Slavic, Germanic, Greek, Indo-Iranian, and Italo-Celtic. The root age of the Indo-European family has been a heavily debated topic since the application of Bayesian phylogenetic methods to lexical cognate data. It was first estimated using phylogenetic methods developed in computational biology (Gray and Atkinson, 2003; Atkinson et al., 2005; Nicholls and Gray, 2008; Ryder and Nicholls, 2011; Bouckaert et al., 2012). These phylogenetic methods employ lexical cognate data (from Swadesh word lists [cf. Table 3]; Swadesh 1952) and external evidence (from archeology and history) regarding both the age of the ancient languages (such as Latin) and the age of the internal subgroups (such as Germanic) to infer the timescale of the Indo-European phylogeny. The work of Gray and colleagues produced root age estimates that supported the Anatolian origin hypothesis (8000–9500 years before present [BP]; Renfrew, 1987) for the Indo-European language family. In contrast, historical linguistics – based on cultural and material vocabulary – points to a Steppe origin of the Indo-European language family, where the root age falls within the range 5500–6500 Years BP (Anthony and Ringe, 2015).

In followup work, Chang et al. (2015) corrected the IELex dataset (Dunn, 2012) – originally compiled by Dyen et al. (1992) – and tested a wide range of models and datasets. Chang et al. (2015) modified the Bayesian phylogenetic inference software BEAST (Drummond et al., 2012) in such a way that the software samples trees that show eight ancient languages – Vedic Sanskrit, Ancient Greek, Latin, Classical Armenian, Old Irish, Old English, Old High German, and Old West Norse – as ancestors of modern descendant languages (cf. Table 1). The results of their analysis showed that the estimated median root age of the Indo-European language family falls within the age range supporting the Steppe origin of the Indo-European language family.

Ancient language	Modern descendants
Vedic Sanskrit	Indo-Aryan languages
Ancient Greek	Modern Greek
Latin	Romance languages
Classical Armenian	Modern Armenian dialects: Adapazar, Eastern Armenian
Old Irish	Irish, Scots Gaelic
Old English	English
Old West Norse	Faroese, Icelandic, Norwegian
Old High German	German, Swiss German, Luxembourgish

Table 1: Ancestry constraints: ancient languages and their descendants employed by Chang et al. (2015).

The phylogenetic dating analysis reported by Bouckaert et al. (2012) and Chang et al. (2015) is based on a coalescent tree prior that employs both the ages of the ancient languages and the internal node ages to infer the dates of all internal nodes (and the root) of a language tree. The coalescent tree prior described in the context of Bayesian phylogenetic inference by Yang (2014: 309–320) is based on the coalescence process studied by Kingman (1982) and is used to model the spread of viruses or alleles in a population of individuals across time.

The coalescent tree prior cannot model the linguistic reality that an ancient language such as Old English is the ancestor of Modern English. It will infer that both Old English and Modern English descended from an unattested linguistic common ancestor. This observation is the basis for the ancestry constrained analyses reported by Chang et al. (2015). The authors found that constraining an ancient language to be the ancestor of modern language(s) infers a reduced age for the root of the Indo-European language family, which supports the Steppe origin hypothesis.

Chang et al. (2015) also observed that the coalescent tree prior without ancestry constraints does not sample trees where an ancient language can be the ancestor of modern language(s). The observation that the coalescent tree prior might not be appropriate for modeling the evolution of the Indo-European family also marks the departure point of the analyses reported in this paper, where I explore the effect of tree priors in Indo-European phylogenetics. All previous phylogenetic studies involving the Indo-European family compare the fit and effect of the age of different substitution models such as Covarion, Stochastic Dollo, and a binary state Generalized Time Reversible model. However, none of the above studies examines the effect of tree priors on dating of the Indo-European language family.

In this paper, I attempt to fill this gap by analyzing all five publicly available datasets (cf. Section 3.1) using an FBD tree prior, a uniform prior, and a constant population size coalescent prior. I perform a Bayes factor analysis similar to Chang et al. (2015) in Section 3.5, and find that the trees inferred with the FBD prior (Stadler, 2010; Heath et al., 2014; Gavryushkina et al., 2014; Zhang et al., 2015) and the uniform tree prior (Ronquist et al., 2012a) support the Steppe origin hypothesis for the Indo-European languages. Finally, the root's median age and 95% highest posterior density ages inferred from the coalescent analysis support an Anatolian origin of the Indo-European languages.

Unlike Bouckaert et al. (2012) and Chang et al. (2015), I do not supply any subgroup constraint information to the phylogenetic program beforehand, but allow the tree inference program to infer the tree topology along with the divergence times of the internal nodes. I find that the Bayesian phylogenetic program infers known subgroups correctly across tree priors. My experiments with the FBD and uniform priors show that ancestry constraints are not necessary to infer support for the Steppe origin of the Indo-European family. I also performed a model comparison based on the Akaike Information Criterion through MCMC (AICM; Baele et al., 2012) and found that both the uniform and FBD priors fit better than the coalescent tree prior.

The rest of the paper is organized as follows. I will motivate the appropriateness of the FBD prior for the Indo-European family diversification scenario and describe other tree priors in Section 2. I will discuss the datasets, substitution model, tree prior settings, Monte Carlo Markov Chain settings, and calculation of Bayes Factor support for the Steppe origin hypothesis vs. the Anatolian origin hypothesis in Section 3. In section 4, I will present the inferred median ages and 95% highest posterior density (HPD) age intervals, Bayes factors, relevance of ancestry constraints, and quality of inferred trees. Section 5 concludes the paper. [1]

2 Tree priors

In this section, I will describe the three different tree priors used in the paper. The coalescent tree prior is presented in Section 2.1. In section 2.2, I will motivate why the FBD tree prior is more suitable than the coalescent tree prior for the Indo-European family. Finally, I describe the uniform tree prior in section 2.3.

2.1 Constant size coalescent prior

The constant population size coalescent tree prior is dependent on the $\theta(= 2Pc)$ parameter, where P is the effective population size and c is the base clock rate. The probability of a tree under this model is $\prod_{j=2}^{n} \frac{2}{\theta} \exp(-\frac{j(j-1)}{\theta} t_j)$, where t_j is the time during which there are j lineages ancestral to the sequences in the data. Both P and c are sampled in this paper. Note that the constant size population prior was also used by Chang et al. (2015: A6,220) to perform an ancestry constrained phylogenetic analysis which supports the Steppe origin hypothesis. To the best of my knowledge, I am not aware of any previous interpretation of a coalescent process in a linguistic scenario. My interpretation when applying the constant size coalescent prior to languages is that the observed languages are lineages from a large haploid population of individual languages, where each language is spoken in a community.[2]

2.2 Birth-Death priors

Birth-Death tree priors are used to model lineage diversification and to date the split event within a phylogeny. The standard birth-death prior of Yang and Rannala (1997) is conditioned on the age of the most recent common ancestor (t_{mrca}) and assumes that birth (λ) and death (μ) rates are constant over time. In this model, all the tips in the tree are extant and do not contain any fossils (cf. Figure 1b). A fossil can be the ancestor of a modern language or can be extinct without leaving any descendants. For instance, Vedic is considered to be the ancestor of all the modern Indo-Aryan languages (cf. Table 1), whereas Hittite or Gothic are languages that died out without leaving any descendant.

The birth-death model described by Yang and Rannala (1997) handles incomplete languages sampling through $\rho = \frac{n}{N}$, where n is the number of languages in the sample and N is the total number of extant languages in the family. The birth-death model estimates the species divergence times on a relative scale. The relative times can be converted into a geological time scale by tying one or more internal nodes to known historical or archaeological evidence. It should be noted that the coalescent process is mathematically different from the birth-death process (Stadler, 2009: 62–63).

In the case of the Indo-European language family, the standard birth-death tree prior of Yang and Rannala (1997) *only* uses internal node calibrations (for instance, the information that the Germanic

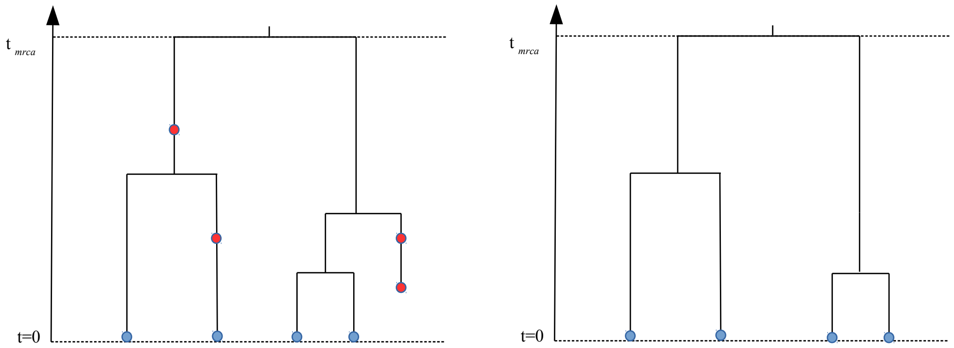
[1] The scripts and the data files used in this paper are available at `https://github.com/PhyloStar/ie-phylo-exps`.

[2] This interpretation is due to Igor Yanovich.

subgroup is about 2200 years old; Chang et al., 2015) to infer the remaining internal nodes' dates. This procedure is known as *node dating* and has been used for inferring the phylogeny of Bantu [3] (Grollemund et al., 2015) and Turkic languages (Hruschka et al., 2015).[4]

The node dating method does not utilize the available lexical cognate information about attested ancient languages that became extinct (e.g. Gothic) or evolved into modern languages (e.g. Latin). However, this method indirectly uses the age information for extinct languages to apply constraints to the internal node ages of a language family. In another argument against node dating, Ronquist et al. (2012a) noted that, if there is more than one fossil in the same language group, then, only the oldest fossil provides the age constraint for the associated internal node. For example, in the case of the Germanic subgroup, there are four fossil languages – Gothic, Old High German, Old English, and Old West Norse – out of which only the age information for Gothic would be used to specify the minimum age of the Germanic subgroup, whereas the rest of the fossil languages cannot provide extra information regarding the age of the Germanic subgroup.

Stadler (2010) proposed an extension to the standard birth-death prior that can handle the placement of ancient languages as tips or as internal nodes (fossils; cf. Fig. 1a). This prior is known as fossilized birth-death (FBD) prior since it can handle both fossil and extant species in a single model. The FBD family of priors can model the linguistic fact that Old English is the ancestor of Modern English. Along with the parameters λ and μ, the FBD prior also features the fossil sampling rate parameter ψ, which is the rate at which fossils are observed along a branch. The FBD tree prior requires only the ages of fossils to infer the root age of a tree; it is more objective than node dating, which requires internal node age constraints that are not directly observed. The standard birth-death prior conditioned on t_{mrca} is a special case of FBD prior with $\psi = 0$ (Stadler, 2010: 401). An example of a fossilized birth-death tree is presented in Fig. 1a. The left tree (1a) in Fig. 1 shows the FBD tree including lineages with sampled extant and fossil languages, whereas the right figure (1b) shows the standard birth-death tree with extant languages.



(a) Fossilized birth-death tree

(b) Birth-death tree

Figure 1: The red dots show fossils and the blue dots show the extant languages (Zhang et al., 2015). (a) FBD tree with fossils as both tips and ancestors of modern languages. (b) The corresponding standard birth-death tree with extant languages. $t = 0$ is the present time, whereas t_{mrca} is the age of the most recent common ancestor.

The probability of a tree under the FBD tree prior is conditioned on t_{mrca} and the nature of extant taxa sampling. In this paper, I assume that the extant taxa are sampled uniformly at random. Unlike

[3] To be precise, the scholars used a pure birth (Yule) process with $\mu = 0, \rho = 1$, a special case of birth-death process, to estimate the divergence times of the internal node splits in the Bantu language family phylogeny.

[4] Hruschka et al. (2015) use cognate sets from an etymological dictionary, where the reflexes within a cognate set need not have the same meaning. This approach is different from the phylogenetic approaches used in this and other papers, where the cognates are root-meaning pairs derived from Swadesh lists (Chang et al., 2015: 201).

Chang et al., who impose ancestry constraints externally, the FBD tree prior can infer the ancestry constraints from the data (if such a signal exists) and do not have to be supplied beforehand. The species sampling probability ρ is determined as the ratio between the number of extant languages in the dataset and the total number of extant Indo-European languages.

The probability of the tree under the FBD model (Stadler, 2010: equation 5) conditioned on x_1 (t_{mrca}) is given below. Here, $n(> 1)$ is the number of extant sampled tips, $m(\geq 0)$ is the number of extinct sampled tips, $k(\geq 0)$ is the number of sampled ancestors with sampled descendants, and y_i is the age of a extinct sampled tip.

$$\frac{\lambda^{n+m-2}\psi^{k+m}}{(1-\hat{p}_0(x_1))^2} p_1(x_1) \prod_{i=1}^{n+m-1} p_1(x_i) \prod_{i=1}^{m} \frac{p_0(y_i)}{p_1(y_i)} \tag{1}$$

Here, $p_0(t)$, $p_1(t)$, c_1, c_2, and $\hat{p}_0(x_1)$ are defined as follows:

- $p_0(t)$ is the probability that an individual present at time t before present has no sampled extinct or extant descendants, which is given as:
  - $p_0(t) = \frac{\lambda+\mu+\psi+c_1 \frac{(\exp(-c_1 t)(1-c_2))-(1+c_2)}{(\exp(-c_1 t)(1-c_2))+(1+c_2)}}{2\lambda}$
- $p_1(t)$ is the probability that an individual present at time t before present has only one sampled extant descendant and no sampled extinct descendants, which is given as:
  - $p_1(t) = \frac{4\rho}{2(1-c_2^2)+\exp(-c_1 t)(1-c_2)^2+\exp(c_1 t)(1+c_2)^2}$
- $\hat{p}_0(x_1) = p_0(t|\psi = 0)$, $c_1 = |\sqrt{(\lambda-\mu-\psi)^2 + 4\lambda\psi}|$, $c_2 = -\frac{\lambda-\mu-2\lambda\rho-\psi}{c_1}$

FBD tree priors have been used for estimating divergence times for datasets with extant and fossil species (Heath et al., 2014; Gavryushkina et al., 2014; Zhang et al., 2015). Since the Indo-European family has both fossils and extant languages, the FBD tree prior that handles attested fossil ancestors is more suitable than the coalescent tree prior that treats fossils as tips. For instance, Tocharian languages became extinct without leaving any modern descendant language, whereas modern Romance languages are the descendants of Latin (an ancient language). Moreover, the data for the Indo-European language family comes from divergent languages and not from a single population. These arguments support the choice of a FBD prior over a coalescent prior for modeling the evolution of the Indo-European language family.

2.3 Uniform tree prior

Similar to the coalescent tree prior, the uniform tree prior (Ronquist et al., 2012a) treats fossils as tips of the tree. However, the uniform tree prior does not make any assumptions regarding the lineage diversification process. The uniform tree prior assumes that the internal nodes' ages are uniformly distributed between tip ages and the root age. The prior probability of a tree under the uniform model is conditioned on $r(t_{mrca})$, which is drawn from a prior distribution h. Under this model, an interior node age is drawn from a uniform distribution with a tip age as the lower bound and the root age as the upper bound. The probability of a tree under the uniform model is proportional to $h(r) \prod_{j=1}^{n-2} \frac{1}{r-t_{j+1}}$, where t_j is the age of a tip j.

3 Methods

In this section, I describe the datasets, prior settings, inference procedure details, and calculation of the Bayes Factor.

3.1 Data

Language	Age Prior	Language	Age Prior
Hittite	3500 – 3600	Old High German$^{\text{A}}$	1000 – 1100
Old Irish$^{\text{A}}$	1100 – 1300	Tocharian B	1200 – 1500
Classical Armenian$^{\text{A}}$	1300 – 1600	Tocharian A	1200 – 1500
Ancient Greek$^{\text{A}}$	2400 – 2500	Lycian	2350 – 2450
Luvian	3275 – 3425	Old Prussian	500 – 600
Vedic Sanskrit$^{\text{A}}$	3000 – 3500	Umbrian	2100 – 2300
Old English$^{\text{A}}$	950 – 1050	Avestan	2450 – 2550
Old Persian	2375 – 2525	Gothic	1625 – 1675
Latin$^{\text{A}}$	2100 – 2200	Old Norse$^{\text{A}}$	750 – 850
Oscan	2100 – 2300	Old Church Slavonic	950 – 1050
Cornish	300 – 400	Sogdian	1200 – 1400

Table 2: Calibration dates for the ancient/medieval languages. All dates are given as years before present (BP). The superscript $^{\text{A}}$ denotes those languages that are assumed to be ancestors of extant languages by Chang et al. (2015).

All five datasets used in this paper – B1, B2, Broad, Medium, and Narrow – are assembled from IELex by Chang et al. (2015).[5] The B1 dataset is derived from Bouckaert et al. (2012) and consists of 207 meanings for 103 languages. The B2 dataset consists of 97 languages and is a subset of the B1 dataset; it is obtained after discarding six languages (Lycian, Oscan, Umbrian, Old Persian, Luvian, and Kurdish) that have attestations for less than 50% of the meanings.

The Broad dataset consists of 94 languages and 197 meaning classes.It has been corrected for cognate judgments in the Indo-Iranian subgroup also contains an extra medieval language, Sogdian, which is not present in B1. Ten meanings that are susceptible to sound symbolism and have poor coverage in terms of the number of languages were also removed from the Broad dataset (Chang et al., 2015: 213). The Medium dataset is a subset of the Broad dataset and is assembled in such a way that the languages and meanings with poor coverage are excluded. It comprises 82 languages and 143 meanings. The Narrow dataset, in turn, is a subset of the Medium dataset and consists of only those modern languages that have an attested ancestor. This selection leaves the Narrow dataset with 52 languages.[6]

3.2 Substitution models

Bayesian phylogenetics originated in evolutionary biology and aims at inferring the evolutionary relationship (trees) between DNA sequences of species. The same method can also be applied to binary (morphological) traits of species (Yang, 2014). Linguistic data is binary trait data, with each cognate class a column in the trait matrix: words that belong to a given cognate class are coded as `1`, else, they are coded as `0`. For example, in the case of German, French, Swedish, and Spanish, the word for *all* in German [alə] and Swedish [ˈalːa] would belong to the same cognate set as English, while French [tu] and Spanish [toðo] belong to a different cognate set. The binary trait matrix for these languages is shown in Table 3 for the two meanings ALL and AND. If a language is missing in a cognate set, then the entry for that language is coded as `?`, and is ignored in the calculation of likelihood using the pruning

[5] One of the reviewers asked why I did not experiment with CoBL database (`http://www.shh.mpg.de/207610/cobldatabase`). The database is not publicly available to perform experiments.

[6] All the datasets are available at `http://muse.jhu.edu/article/576999/file/supp02.zip`.

algorithm (Felsenstein, 2004: 255). I used a Generalized Time Reversible model (equivalent to a F81 model in the case of binary traits) with ascertainment bias correction (Felsenstein, 1992; Lewis, 2001) for all unobserved **0** columns. The rate variation across sites is modeled using a discrete Gamma model with four rate categories (Yang, 1994), where the shape parameter of the Gamma distribution is drawn from an exponential prior with mean 1.

Language	ALL	AND	...
English	ɔːl^1	ænd^1	...
German	alə^1	ʊnt^1	...
French	tu^2	e^2	...
Spanish	toðo^2	i^2	...
Swedish	ˈalːa^1	ɔkː^3	...

(a) Forms and cognate classes

Language	ALL		AND		
English	1	0	1	0	0
German	1	0	1	0	0
French	0	1	0	1	0
Spanish	0	1	0	1	0
Swedish	1	0	0	0	1

(b) Binary matrix

Table 3: Excerpt from meaning list showing cognate classes (a) and the binary cognate matrix (b) for meanings ALL and AND in five languages. The superscript indicates words that are cognate.

3.3 Tree prior settings

In this paper, I assume that the extant languages are randomly sampled. The FBD tree prior is dependent on the number of extant languages in the sample. I estimated the number of extant Indo-European languages (400) from Glottolog (Nordhoff and Hammarström, 2012), and set the ρ parameter accordingly for each dataset. For the FBD prior, the net diversification rate $d(=\lambda-\mu)$ is drawn from an exponential prior with mean 1, the relative extinction rate (turnover) $r(=\mu/\lambda)$ is drawn from a Beta(1,1) prior, and the fossil sampling probability $f(=\psi/(\psi+\mu))$ is also drawn from a Beta(1,1) prior.

I draw the root age from a uniform distribution bounded between 4000 and 25000 years in the case of the FBD and uniform priors. The root age's upper bound is fixed at 25000 years since this age is more than double the upper bound of the age limit for the Anatolian origin hypothesis. In fact, none of the inferred trees' root ages are even close to 25000 years. The coalescent prior, as implemented in MrBayes, is not conditioned on t_{mrca}. All the fossils' age priors were drawn from uniform distributions, whose age ranges are given in Table 2.

In the case of the coalescent prior, the population parameter P is drawn from a Gamma distribution with shape parameter 1 and rate parameter 0.01.[7] The base clock rate c is drawn from an exponential prior with mean 10^{-4}. In all the analyses, I use a Independent Gamma Rate model (Lepage et al., 2007), with each branch rate drawn from a Gamma distribution with mean 1.0 and variance σ^2_{IG}/b_j, where b_j – the branch length of a branch j – is computed as the product of geological (or calendar) time t_j and c. σ^2_{IG} is the independent gamma rate model's variance parameter that is drawn from an exponential prior with mean 0.005. I do not employ topology constraints and allow the software to infer the Indo-European phylogeny from the data, along with the time scale.

3.4 Markov chain Monte Carlo sampling

All experiments were conducted using MrBayes software.[8] I carried out two independent runs (each run consisted of one cold chain and two hot chains) and verified that the average standard deviation of

[7] I discovered a bug in the MrBayes implementation with the coalescent prior that was calculating the Metropolis-Hastings ratio incorrectly. My implementation is already made available here: `https://github.com/PhyloStar/mrbayes-coal`.

[8] Available at `http://mrbayes.sourceforge.net/`.

split frequencies (Ronquist et al., 2012b) between both runs was less than 0.01. I ran all the analyses for 20–80 million states and sampled every 1000^{th} state to reduce auto-correlation between the sampled states. For each dataset, I discarded the initial 25% of the states as burn-in and generated a 50% majority rule consensus tree[9] from the remaining 75% of the states (Felsenstein, 2004: chapter 30).[10]

3.5 Evaluating Steppe vs. Anatolian hypothesis

For each dataset, I ran the MrBayes software twice: once without cognate data to generate a prior sample of trees, and once with cognate data to generate a posterior sample of trees. Then, I used the Bayes Factor (BF) formulation from Chang et al. (2015) to calculate the support for the Anatolian (A) and Steppe (S) hypotheses. Given data D, the Bayes factor $K_{S/A}$ is calculated as follows:

$$\frac{\mathbb{P}(D|t_R \in \Omega_S)}{\mathbb{P}(D|t_R \in \Omega_A)} \tag{2}$$

where $\Omega_S \in [5500, 6500]$ and $\Omega_A \in [8000, 9500]$ represents the range of Steppe and Anatolian ages, respectively, t_R denotes the root age of a tree, which is t_{mrca} in the case of the FBD prior. The numerator and denominator in equation 2 are computed as follows:

$$\frac{Pr\{t_R \in \Omega_S|D\}}{Pr\{t_R \in \Omega_S\}} / \frac{Pr\{t_R \in \Omega_A|D\}}{Pr\{t_R \in \Omega_A\}} \tag{3}$$

The numerators $Pr\{t_R \in \Omega_S|D\}, Pr\{t_R \in \Omega_A|D\}$ in equation 3 correspond to the fraction of trees in the posterior sample for which $t_R \in \Omega_S$ and $t_R \in \Omega_A$, that is, those trees in the posterior sample whose roots fall within the time frames for the Steppe and Anatolian hypotheses, respectively. The denominators $Pr\{t_R \in \Omega_S\}, Pr\{t_R \in \Omega_A\}$ correspond to the fraction of trees in the prior sample for which this is the case. Following the interpretation of Bayes Factor by Kass and Raftery (1995), the support for the Steppe origin hypothesis is very strong if $K_{S/A} > 150$, strong if $20 < K_{S/A} < 150$, positive if $3 < K_{S/A} < 20$, not worth more than a bare mention (*neutral*) if $1 < K_{S/A} < 3$ and negative if $K_{S/A} < 1$.

4 Results

In this section, I present and discuss the root's median age and 95% HPD age intervals, fit of tree prior, Bayes Factor support for the Steppe vs. the Anatolian hypotheses, comparison of subgroups' inferred dates with expert dates, relevance of clade constraints, and ancestry constraints.

4.1 Median and 95% HPD ages

Table 4 shows the HPD intervals and median root ages for all dataset and tree prior combinations. None of the reported HPD age intervals lie completely within the Steppe or the Anatolian age interval. The lower bounds of HPD ages in the case of the FBD and uniform priors fall within the Steppe interval, whereas the lower bound of the coalescent prior's HPD interval falls beyond the Steppe age interval. In the case of the NARROW and MEDIUM datasets, the root age is further reduced to 6826 and 6845 years, respectively, in the case of FBD prior. The median ages inferred by FBD prior belong neither to the Steppe hypothesis interval nor to the Anatolian hypothesis interval for all datasets. The median ages inferred by the uniform prior for the BROAD, MEDIUM, and NARROW datasets lie within the range of

[9] A 50% majority consensus tree is a summary tree that consists of only those clades that occur in more than 50% of the post burn-in sample of trees.

[10] I present the inferred phylogenies, posterior support and HPD intervals of the internal nodels for all the tree priors and datasets in the appendix.

Dataset	95% HPD			Median Age		
	FBD	Coalescent	Uniform	FBD	Coalescent	Uniform
B1	6244–8766	8370–11695	5760–8115	7512	9821	6789
B2	6150–8430	7590–10913	5536–7986	7177	9133	6738
BROAD	5591–7585	6654–9327	5073–6947	6551	7984	5935
MEDIUM	5942–7921	7070–9818	5395–7392	6845	8345	6339
NARROW	5790–7984	6826–9791	5423–7646	6826	8228	6462

Table 4: Columns 2–4 show the 95% Highest Posterior Density (HPD) and columns 5–7 show the median ages (in years before present) of the root node from the consensus tree for each dataset and a tree prior.

the Steppe interval. All priors infer median ages that lie beyond the Steppe interval in the case of the B1 and B2 datasets. The coalescent prior infers root ages that lie within the Antolian hypothesis in the case of all the datasets except B1. Across all priors, the median root ages decrease when the datasets are corrected for errors. The descreasing trend in the median ages is similar to the trend observed in Chang et al. (2015).

Why does the BROAD dataset yield younger ages? Chang et al. (2015) argue that sparsely attested languages can influence the chronology estimates. They note that the ascertainment bias correction to the likelihood calculation (Felsenstein, 1992) accounts for unobserved cognate sets that are not attested in the data, but does not account for the missing entries in a dataset. For example, if 50% of the data is missing for a language, then the ascertainment bias correction does not account for these missing 50%. If there are x unique cognate sets in the observed 50% of the data, then, there is a possibility that the unobserved 50% of the data also contains x unique cognate sets that do not enter the likelihood calculation.

The likelihood calculation would only consider the observed x unique cognate sets, therefore underestimating the true number of unique cognate sets for a language in a dataset. For this reason, a language with a higher number of missing entries is treated as more conservative (in other words, is inferred to have undergone a lesser number of character changes) than it should be. This is particularly true for languages such as Hittite and Tocharian A & B, which have about 11.95% and 32.74% missing entries, respectively, in the case of the BROAD dataset as compared to 1.92% and 2.13%, respectively, in the case of the MEDIUM dataset. Since both the Hittite and Tocharian doculects are very close to the root of the Indo-European tree, this underestimation of the number of unique cognate sets leads to a shorter branch length, which causes the median root age to be younger. Both the coalescent and FBD tree priors infer a younger age for the BROAD dataset than for the MEDIUM and NARROW datasets.

Why does the B2 dataset yield younger ages? The B1 dataset features six sparsely attested languages – Lycian, Oscan, Umbrian, Old Persian, Luvian, and Kurdish – for which more than 50% of the meanings are unattested. As explained in the previous paragraph, inclusion of sparsely attested languages causes the Bayesian inference program to underestimate the root age. The opposite happens when a language has more number of unique cognates than it should have. This is the case for Luvian, where 33% of the attested cognate sets are erroneously coded as unique cognate sets even though they are cognate with either Hittite or Lycian. This erroneous coding causes the Bayesian software to treat Luvian, which is one internal node away from the root node, as having evolved more and posits longer branches, thereby pushing the root age of the tree away from the Steppe age interval. The B2 dataset excludes the six sparsely attested languages including erroneously coded Luvian, which leads to shortening of the median root age in the posterior sample. This effect is clearly observed with both the median root age and the 95% HPD age range in the B2 dataset. When the FBD tree prior is applied to this dataset, the median root age is pushed 400 years downwards towards the Steppe hypothesis. The coalescent prior likewise infers a younger median age for the B2 dataset than for the B1 dataset, whereas the uniform prior is not influenced by the six sparsely attested languages.

4.2 Which tree prior is the best?

Tree Prior	B1	B2	BROAD	MEDIUM	NARROW
Uniform Prior	**94002.748**	90299.551	**89269.61**	50769.888	**32162.117**
FBD Prior	94005.297	**90297.721**	89270.359	**50764.79**	32163.007
Coalescent Prior	94117.099	90396.491	89374.335	50917.074	32241.019

Table 5: AICM values for each of the datasets. The lower the value, the better the model's fit to the data. The best-fitting model's AICM value is shown in bold.

I determine the best model through Akaike Information Criterion through MCMC (AICM; Baele et al., 2012). It has to be noted that Bouckaert et al. (2012) employ both harmonic mean and AICM to perform model comparison. In this paper, I only use AICM, since it is more accurate than the harmonic mean, which is unstable. On the other hand, methods such as stepping stone sampling (Xie et al., 2010) and thermodynamic integration (Lartillot and Philippe, 2006) used to estimate marginal likelihood are more accurate than AICM, but are computationally intensive and require at most K times (usually set to 10) the computation as the original MCMC runs (Yang, 2014: 258–259).

The AICM values for each dataset and tree prior are presented in Table 5. The results show that the uniform tree prior fits best for the B1, BROAD, and NARROW datasets. The difference between the AICM values of the uniform and FBD priors is almost negligible in the case of the BROAD and NARROW datasets. The coalescent prior shows the highest AICM value by a large margin when compared to the FBD and uniform priors. Since the uniform tree prior has fewer parameters than the FBD prior, I suggest that any future phylogenetic experiment *should test the uniform tree prior as a baseline* before testing more parameter-rich priors such as FBD or coalescent priors.

4.3 Bayes factor for Steppe vs. Anatolia

Dataset	FBD	Coalescent	Uniform
B1	0.138 (Negative)	**	67.043 (Strong)
B2	1.015 (Neutral)	**	1022.968 (Very Strong)
BROAD	88.624 (Strong)	*	6728.994 (Very Strong)
MEDIUM	18.536 (Positive)	**	113.968 (Strong)
NARROW	16.55 (Positive)	*	27.549 (Strong)

Table 6: Bayes factor support for the Steppe origin across different datasets and tree priors. * represents an entry where there is no tree in the prior sample whose root age that falls within the Steppe range. ** indicates those datasets that do not have a posterior and prior root age within the Steppe range.

I present the results of the Bayes factor (BF) analysis in Table 6. In the case of the FBD prior, BF results support the Steppe origin hypothesis for all the datasets, except for the B1 dataset. The corrected datasets clearly support the Steppe hypothesis *positively* in terms of Bayes factor in the case of FBD prior. In the case of the uniform prior, all the datasets *support* the Steppe origin hypothesis over the Anatolian origin hypothesis. In the case of the coalescent prior, the Bayes factor was not possible to calculate since there is no tree in either the prior or the posterior sample whose root age falls within the age range of the Steppe hypothesis. Overall, the interpretation of the strength of the Bayes factor analysis suggests that appropriate tree priors and corrected datasets *support* the Steppe origin hypothesis of the Indo-European language family.

4.4 Internal node ages

In this subsection, for each dataset, I compare the inferred dates for the language subgroups with the historically attested dates given in Table 7. The uniform tree prior, on average, overestimates the ages for all the datasets, except for the NARROW dataset. The predicted ages from the uniform tree prior come closest to the historical ages in the case of the MEDIAN dataset. In contrast, Chang et al. present younger ages for both the NARROW (100 years on an average) and MEDIUM datasets (330 ± 165 years).

Subgroup	Historical Age	B1	B2	BROAD	MEDIUM	NARROW
Germanic	2250	2876 [2286-3572]	2816 [2256-3458]	2615 [2147-3166]	2449 [2031-2935]	2334 [1943-2807]
Romance	1750	2987 [2400-3629]	2149 [1628-2714]	1980 [1515-2493]	1841 [1401-2345]	1736 [1309-2248]
Scandinavian	1500	1523 [1127-2016]	1469 [1102-1906]	1340 [1024-1697]	1164 [898-1477]	–
Slavic	1500	1860 [1401-2423]	1822 [1378-2309]	1647 [1301-2069]	1575 [1226-1972]	–
East Baltic	1300	1584 [914-2356]	1561 [936-2265]	1465 [891-2086]	1460 [892-2115]	–
British Celtic	1250	1732 [1105-2402]	1687 [1137-2343]	1537 [1024-2093]	1450 [955-2011]	–
Modern Irish/Scots Gaelic	1050	1058 [530-1615]	1052 [589-1620]	967 [523-1442]	834 [451-1260]	829 [442-1290]
Persian-Tajik	750	882 [424-1412]	842 [386-1360]	819 [409-1250]	704 [336-1098]	–
Average difference		-394	-256	-127	-15.875	50.33

Table 7: The second column shows the ages of each language subgroup based on historical events. The rest of the columns show the uniform prior's median ages (in years before present) and 95% HPD age intervals (in square brackets) across the five datasets. The last row shows the average difference between historical ages and predicted ages. The historical ages are obtained from Chang et al. (2015: 226). The East Baltic group consists of Lithuanian and Latvian. The British Celtic group consists of Cornish, Breton, and Welsh. The Romance group consists of all the Romance languages except Latin.

4.5 Relevance of clade constraints

Both Bouckaert et al. (2012) and Chang et al. (2015) constrain the topologies in tree search through clade constraints. For instance, a Germanic clade constraint would mean that the Bayesian software would only sample those trees that place all the Germanic languages under a single node. However, the two teams of authors do not follow the same set of topological constraints when inferring the dates for the Indo-European language family. Chang et al. (2015) apply a stricter set of constraints – derived from the linguistic knowledge of the Indo-European language family – than those of Bouckaert et al. (2012). In this paper, I do not employ any clade constraints and allow the software to automatically infer the tree topology from the datasets.[11]

I present the majority rule consensus tree inferred using the uniform prior for the BROAD dataset in Fig. 2. The majority rule consensus tree retrieves the well-established language subgroups such as Balto-Slavic, Greek, Indo-Iranian, Germanic, and Italo-Celtic correctly. In fact, I observe that all consensus trees (cf. A appendix) retreive the subgroups correctly, without the necessity of supplying this information as constraints to the phylogenetic software.

[11] I note that the clade constraint information is derived from historical linguistics research that is limited to language families such as Indo-European, Dravidian, Uralic, Austronesian, and Sino-Tibetan with long tradition of classical comparative linguistic research (Campbell and Poser, 2008).

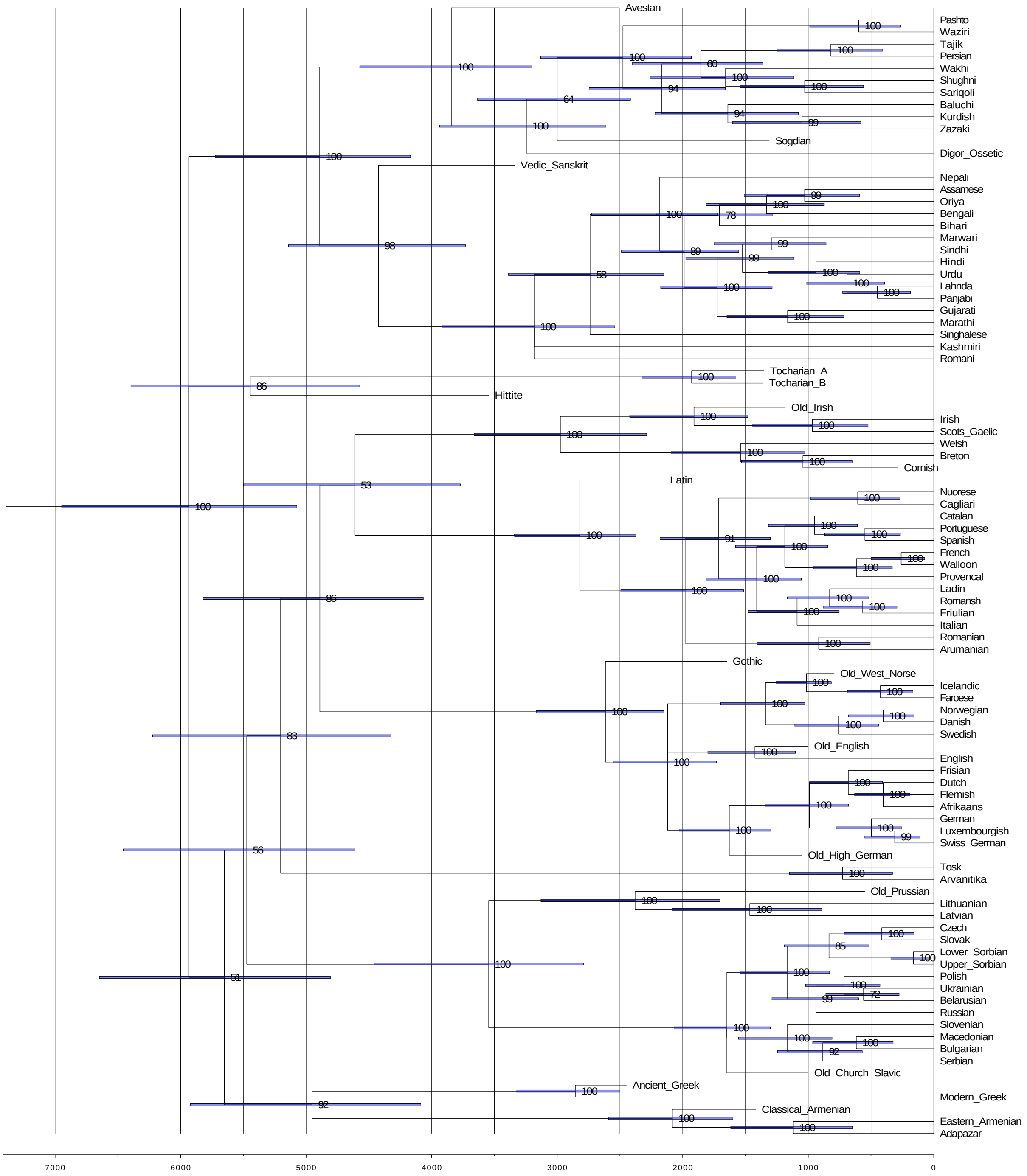


Figure 2: The majority-rule consensus tree inferred using uniform prior for the BROAD dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

Position of Anatolian and Tocharian languages There is a general consensus among the Indo-European scholars that the Anatolian language group which includes Hittite was the first branch to split from the language family tree at the Proto-Indo-European stage, after which the Tocharian language group was the second to split off from the post-Anatolian Indo-European languages (Ringe et al., 2002).

In fact, Chang et al. supply this linguistic knowledge as two constraints to the Bayesian software: they define a Nuclear Indo-European group consisting of all the non-Anatolian languages, and Inner Indo-European group consisting of all the Nuclear Indo-European languages excluding the Tocharian languages. In my study, I observe that the majority consensus trees constructed from the analyses inferred with the uniform tree prior always group both the Anatolian and Tocharian languages as distinct subgroups unified under the same internal node, which is directly connected to the root node. This is also true in the case of the majority consensus tree inferred when the coalescent tree prior is applied to the B2 dataset. The majority consensus trees constructed from the FBD tree prior's analyses always show that the Anatolian languages were the first to split off, followed by the branching of the Tocharian languages. This observation also holds for the the majority consensus trees inferred with colaescent tree priors applied to B1, BROAD, MEDIUM, and NARROW datasets.

In conclusion, the majority consensus trees suggest that the well-established Indo-European subgroups can be inferred directly, and need not be supplied beforehand. The exact placement of these subgroups with respect to each other within the Inner Indo-European clade is a topic of research among scholars and has yet to be determined to full satisfaction (Anthony and Ringe, 2015).

4.6 Relevance of ancestry constraints

Chang et al. (2015) introduced ancestry constraints into their phylogenetic analysis, which, then, supported the Steppe origin hypothesis. The application of the FBD prior can be used to verify if the ancestry constraints can be inferred from the data. The FBD prior can infer whether an ancient language is an ancestral language or a tip in the tree. However, the majority rule consensus trees inferred from all the datasets using the FBD tree prior do not show any support for the ancestry relationships enforced as constraints by Chang et al. (2015). I examined the log files of the MCMC runs and found that the MCMC proposal move (`delete-branch`) in MrBayes, which supports the placement of an ancient language as an internal node, was never accepted during MCMC sampling. At least based on trees inferred from lexical datasets, I conclude that the FBD prior does not infer any of the ancestry relations employed by Chang et al. (2015).

5 Conclusion

In this paper, I addressed the question of the effect of tree priors in Bayesian phylogenetic analysis and found the following.

- My model comparison results suggest that both the uniform and FBD priors show better fit to the datasets of the Indo-European language family than the coalescent prior. Therefore, based on the Bayes factor analysis, I conclude that the Steppe hypothesis is supported by the FBD and Uniform priors for majority of the datasets.
- The FBD tree prior does not infer any ancestry relation from any of the datasets, suggesting that the lexical datasets used in the paper does not contain a signal for ancestry relations.
- I also observe that the Bayesian inference program can infer well-established subgroups correctly from the data. This information need not be supplied beforehand.
- Finally, the experiments reported in the paper suggest that the right tree priors and corrected cognacy judgments are important for estimating the phylogeny and the age of the Indo-European language family.

Acknowledgments

I thank Remco Bouckaert, Johannes Wahle, Armin Buch, Johannes Dellert, Marisa Köllner, Roland Mühlenbernd, and Vijayaditya Peddinti for all the comments and discussions that improved the paper. This paper would not have been possible without the continuous support of Igor Yanovich, Søren Wichmann, Chris Bentz, Gerhard Jäger, Johann-Mattis List, Richard Johansson, Lilja Øvrelid, and Aparna Subhakari. Finally, I thank the anonymous reviewers for all the comments that helped improved the paper further. One of the reviewers provided extensive comments regarding the models and results, from which the final version benefited considerably. All remaining errors are mine. The author is supported by BIGMED and ERC Advanced Grant 324246 EVOLAEMP, which is gratefully acknowledged.

References

Anthony, David W. and Don Ringe. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1(1): 199–219. doi:10.1146/annurev-linguist-030514-124812.

Atkinson, Quentin, Geoff Nicholls, David Welch, and Russell Gray. 2005. From words to dates: Water into wine, mathemagic or phylogenetic inference? *Transactions of the Philological Society* 103(2): 193–219.

Baele, Guy, Philippe Lemey, Trevor Bedford, Andrew Rambaut, Marc A Suchard, and Alexander V Alekseyenko. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* 29(9): 2157–2167.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, and Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097): 957–960.

Campbell, Lyle and William J. Poser. 2008. *Language classification: History and Method*. Cambridge University Press.

Chang, Will, Chundra Cathcart, David Hall, and Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91(1): 194–244.

Drummond, Alexei J, Marc A Suchard, Dong Xie, and Andrew Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29(8): 1969–1973.

Dunn, Michael. 2012. Indo-European lexical cognacy database (IELex). URL `http://ielex.mpi.nl/`.

Dyen, Isidore, Joseph B. Kruskal, and Paul Black. 1992. An Indo-European classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society* 82(5): 1–132.

Felsenstein, Joseph. 1992. Phylogenies from restriction sites: A maximum-likelihood approach. *Evolution* 46(1): 159–173.

Felsenstein, Joseph. 2004. *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates.

Gavryushkina, Alexandra, David Welch, Tanja Stadler, and Alexei J Drummond. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Computational Biology* 10(12): e1003,919.

Gray, Russell D. and Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426(6965): 435–439.

Grollemund, Rebecca, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti, and 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): 13,296–13,301.

Heath, Tracy A, John P Huelsenbeck, and Tanja Stadler. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences* 111(29): E2957–E2966.

Hruschka, Daniel J, Simon Branford, Eric D Smith, Jon Wilkins, Andrew Meade, Mark Pagel, and Tanmoy Bhattacharya. 2015. Detecting regular sound changes in linguistics as events of concerted evolution. *Current Biology* 25(1): 1–9.

Kass, Robert E and Adrian E Raftery. 1995. Bayes Factors. *Journal of the American Statistical Association* 90(430): 773–795.

Kingman, John Frank Charles. 1982. The coalescent. *Stochastic processes and their applications* 13(3): 235–248.

Lartillot, Nicolas and Hervé Philippe. 2006. Computing Bayes Factors Using Thermodynamic Integration. *Systematic Biology* 55(2): 195–207. doi:10.1080/10635150500433722. URL `http://dx.doi.org/10.1080/10635150500433722`.

Lepage, Thomas, David Bryant, Hervé Philippe, and Nicolas Lartillot. 2007. A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution* 24(12): 2669–2680.

Lewis, Paul O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50(6): 913–925.

Nicholls, Geoff K and Russell D Gray. 2008. Dated ancestral trees from binary trait data and their application to the diversification of languages. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70(3): 545–566.

Nordhoff, Sebastian and Harald Hammarström. 2012. Glottolog/Langdoc: Increasing the visibility of grey literature for low-density languages. In *Language Resources and Evaluation Conference*, 3289–3294.

Renfrew, Colin. 1987. *Archaeology and language : The puzzle of Indo-European origins*. London : Cape.

Ringe, Don, Tandy Warnow, and Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100(1): 59–129.

Ronquist, Fredrik, Seraina Klopfstein, Lars Vilhelmsen, Susanne Schulmeister, Debra L Murray, and Alexandr P Rasnitsyn. 2012a. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology* 61(6): 973–999.

Ronquist, Fredrik, Maxim Teslenko, Paul van der Mark, Daniel L Ayres, Aaron Darling, Sebastian Höhna, Bret Larget, Liang Liu, Marc A Suchard, and John P Huelsenbeck. 2012b. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61(3): 539–542.

Ryder, Robin J and Geoff K Nicholls. 2011. Missing data in a stochastic Dollo model for binary trait data, and its application to the dating of Proto-Indo-European. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 60(1): 71–92.

Stadler, Tanja. 2009. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* 261(1): 58–66.

Stadler, Tanja. 2010. Sampling-through-time in birth–death trees. *Journal of Theoretical Biology* 267(3): 396–404.

Swadesh, Morris. 1952. Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96(4): 452–463.

Xie, Wangang, Paul O Lewis, Yu Fan, Lynn Kuo, and Ming-Hui Chen. 2010. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* 60(2): 150–160.

Yang, Ziheng. 1994. Estimating the pattern of nucleotide substitution. *Journal of Molecular Evolution* 39(1): 105–111.

Yang, Ziheng. 2014. *Molecular Evolution: A Statistical Approach*. Oxford: Oxford University Press.

Yang, Ziheng and Bruce Rannala. 1997. Bayesian phylogenetic inference using DNA sequences: A Markov chain Monte Carlo method. *Molecular Biology and Evolution* 14(7): 717–724.

Zhang, Chi, Tanja Stadler, Seraina Klopfstein, Tracy A Heath, and Fredrik Ronquist. 2015. Total-evidence dating under the fossilized birth–death process. *Systematic Biology* 65(2): 228–249.

A Coalescent Prior

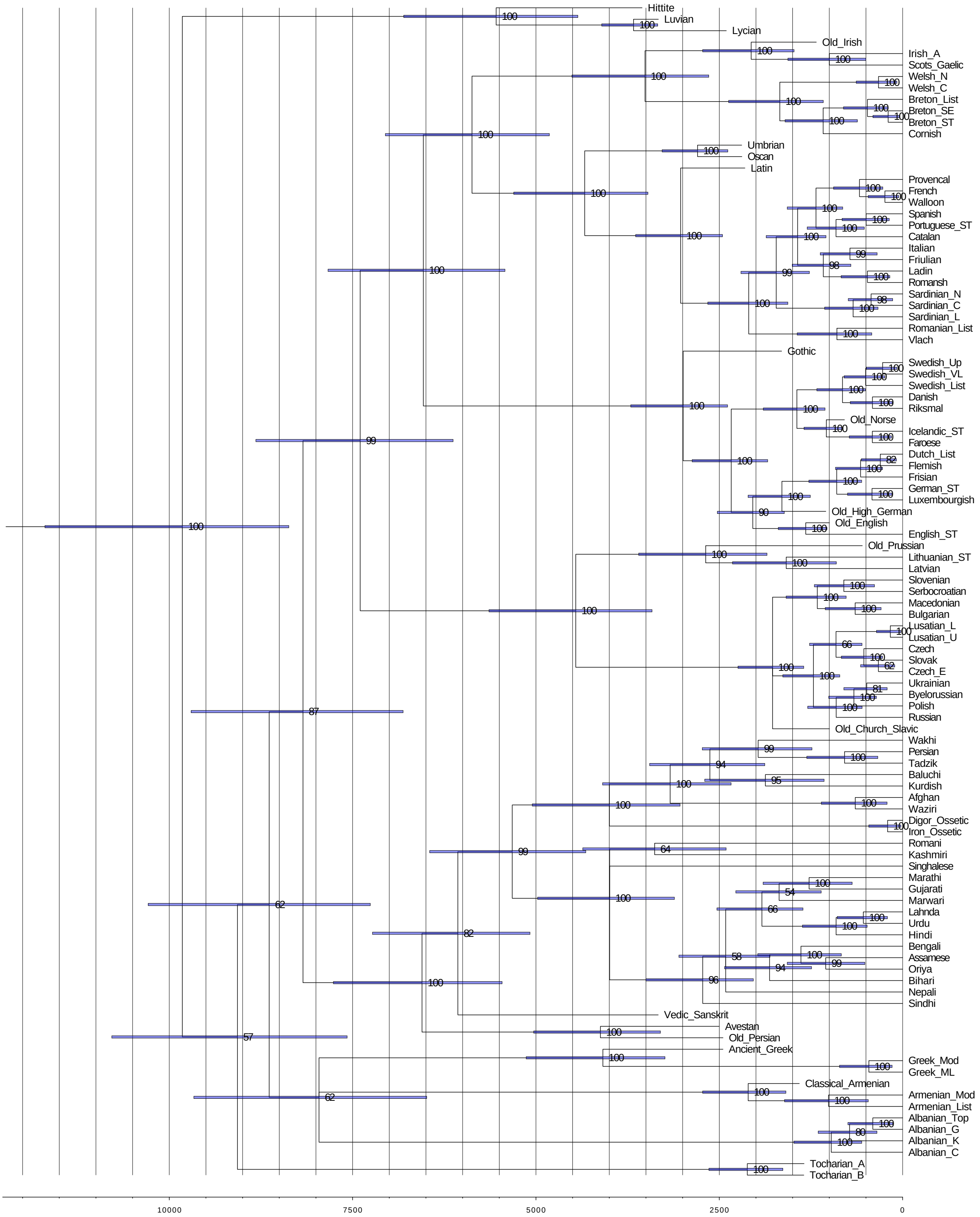


Figure 3: The majority-rule consensus tree for the B1 dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

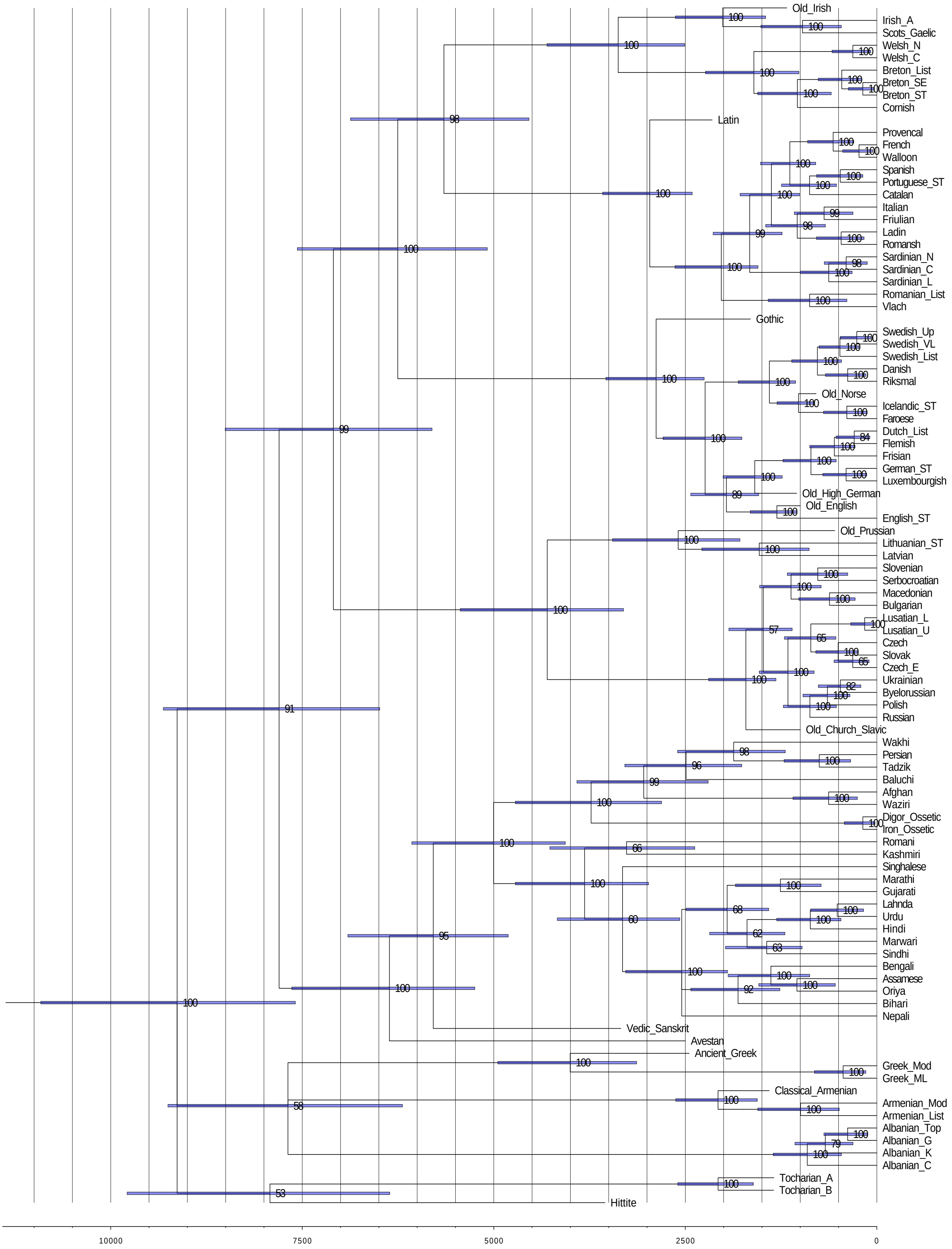


Figure 4: The majority-rule consensus tree for the B2 dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

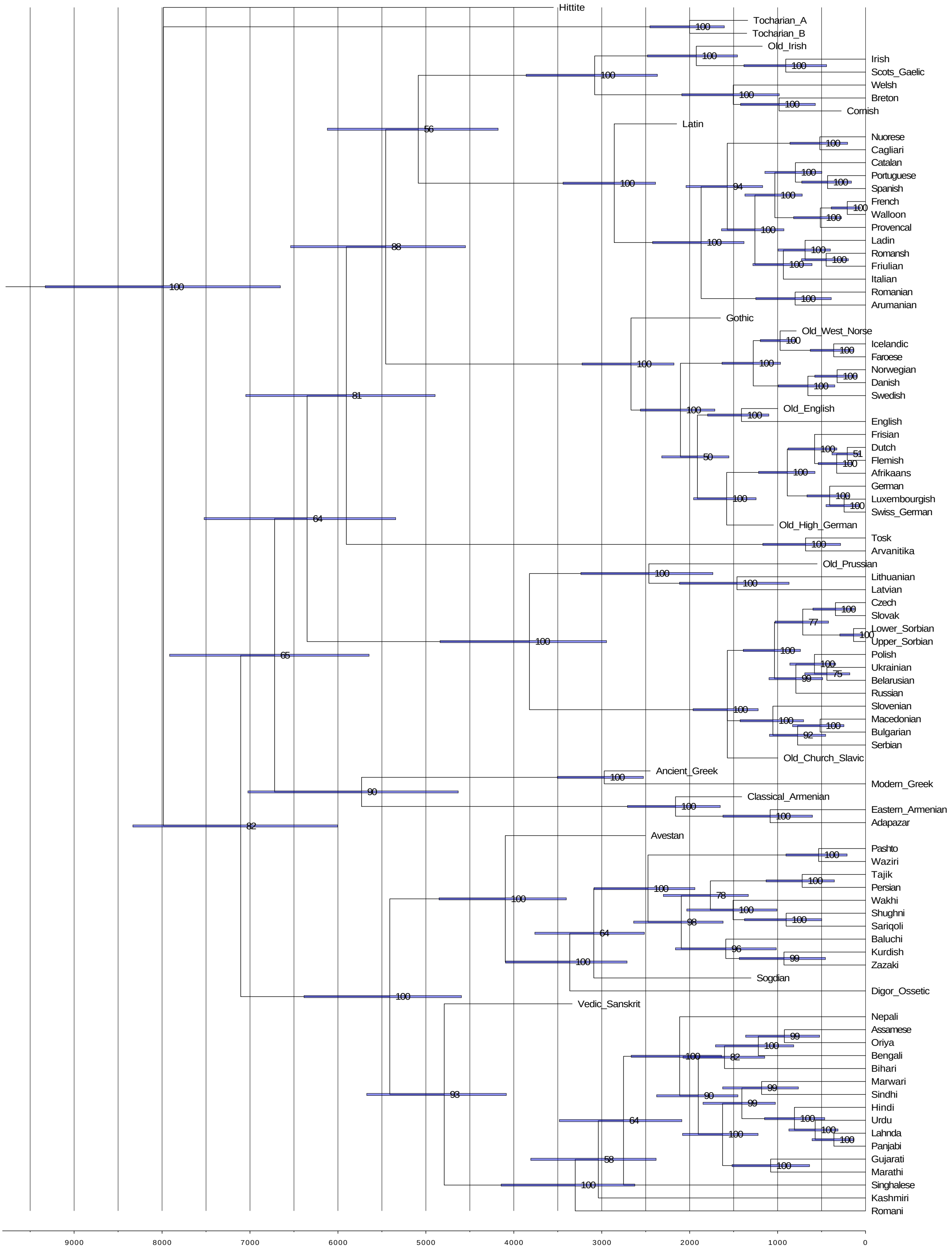


Figure 5: The majority-rule consensus tree for the BROAD dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

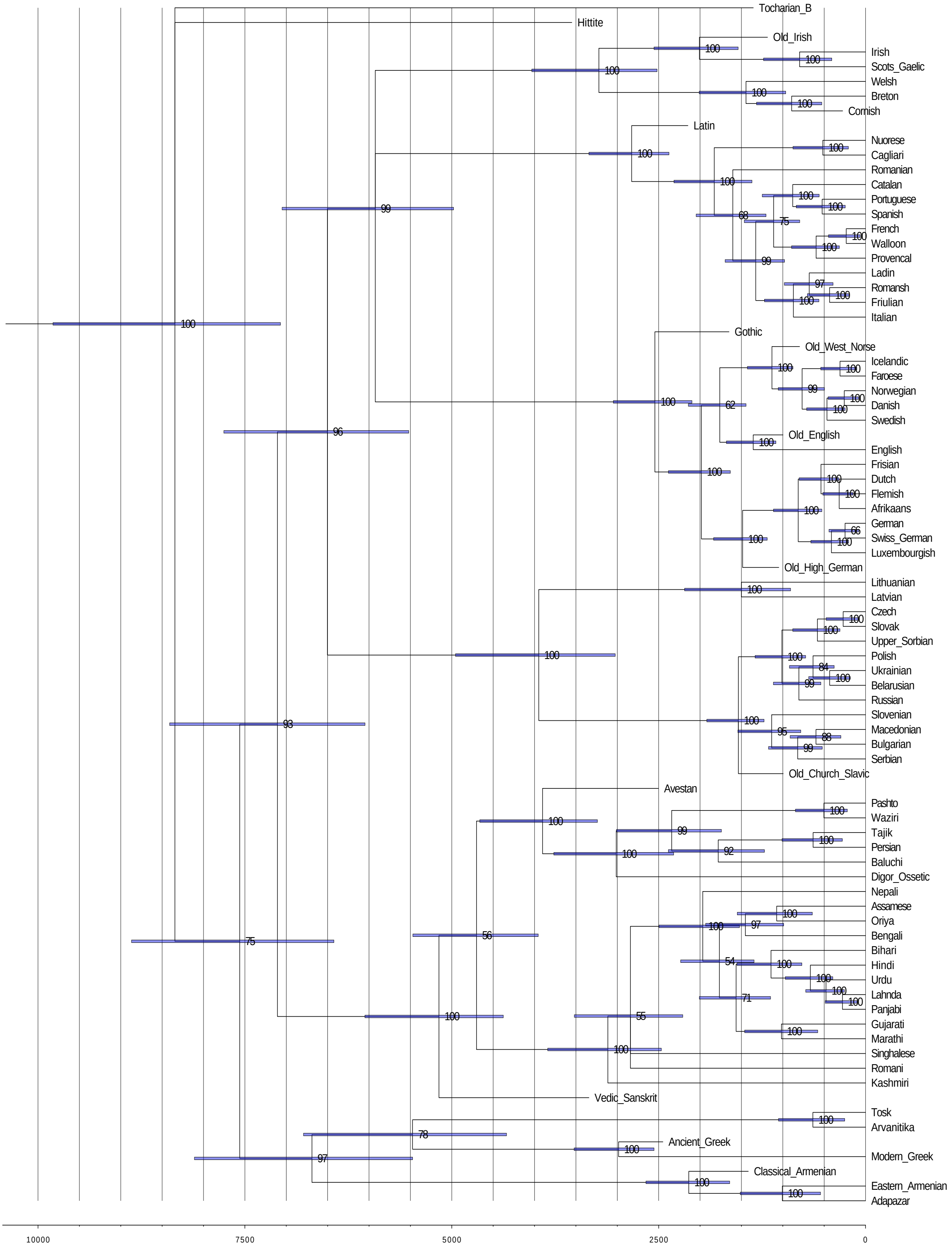


Figure 6: The majority-rule consensus tree for the MEDIUM dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

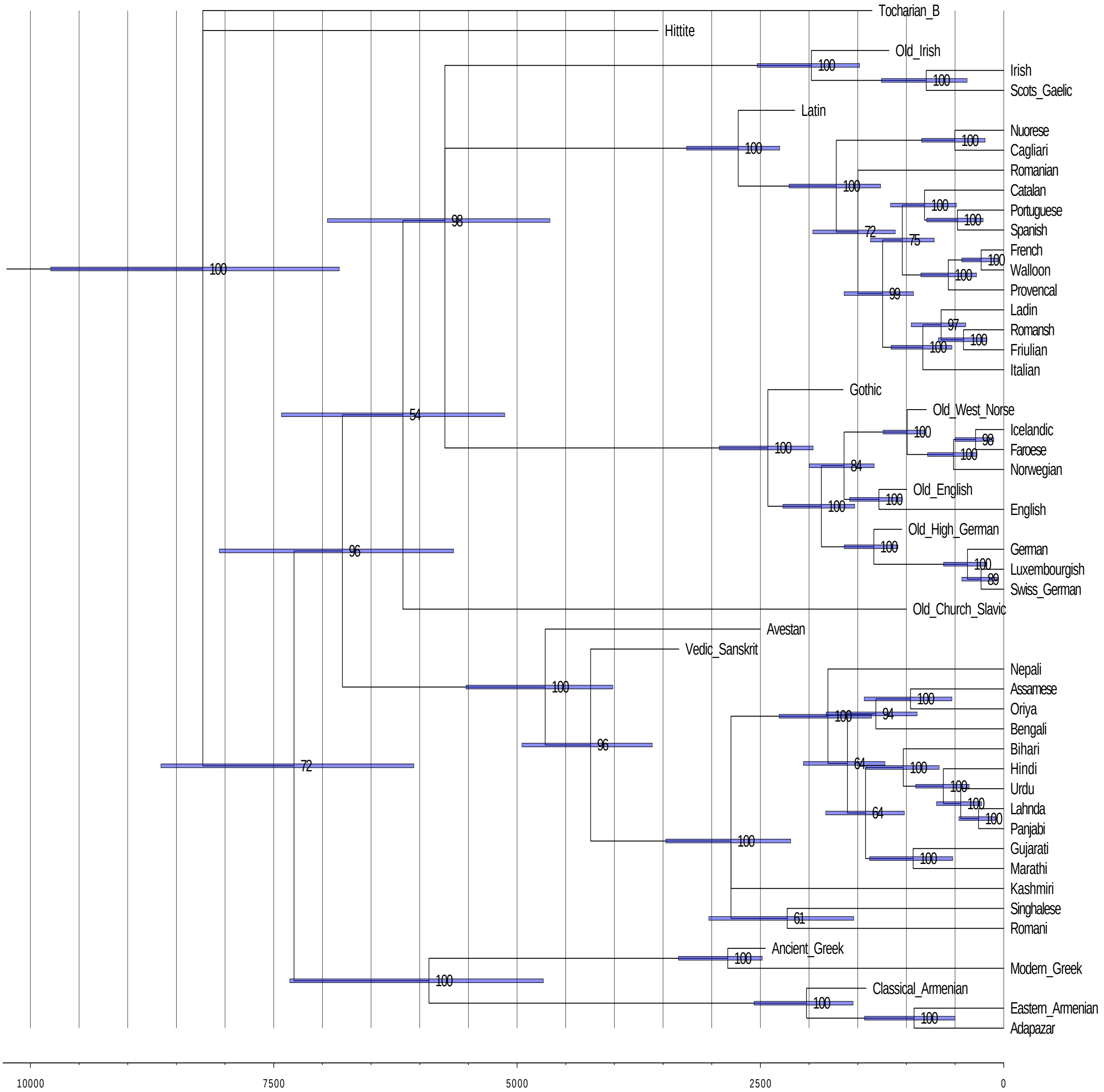


Figure 7: The majority-rule consensus tree for the NARROW dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

B FBD Prior

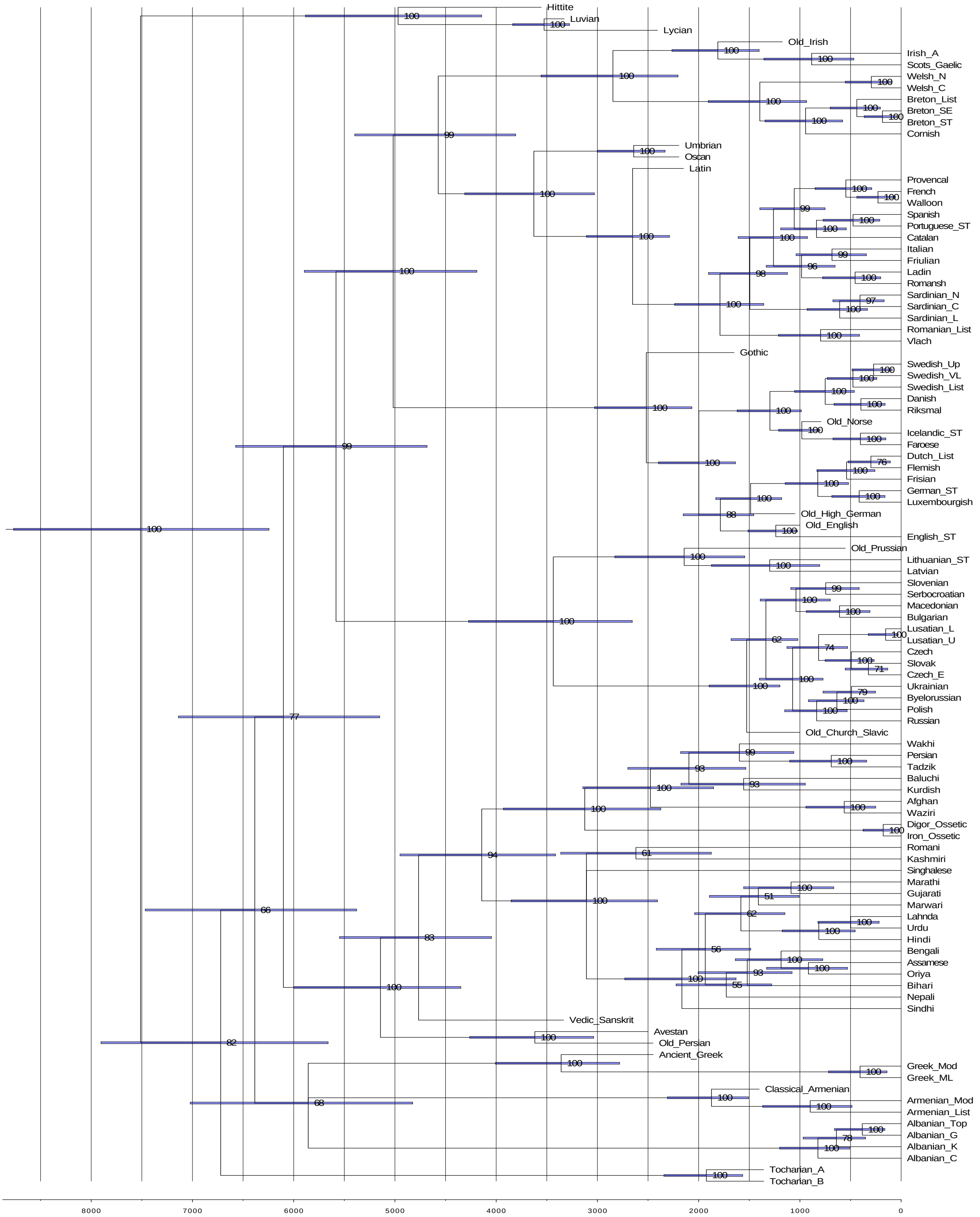


Figure 8: The majority-rule consensus tree for the B1 dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

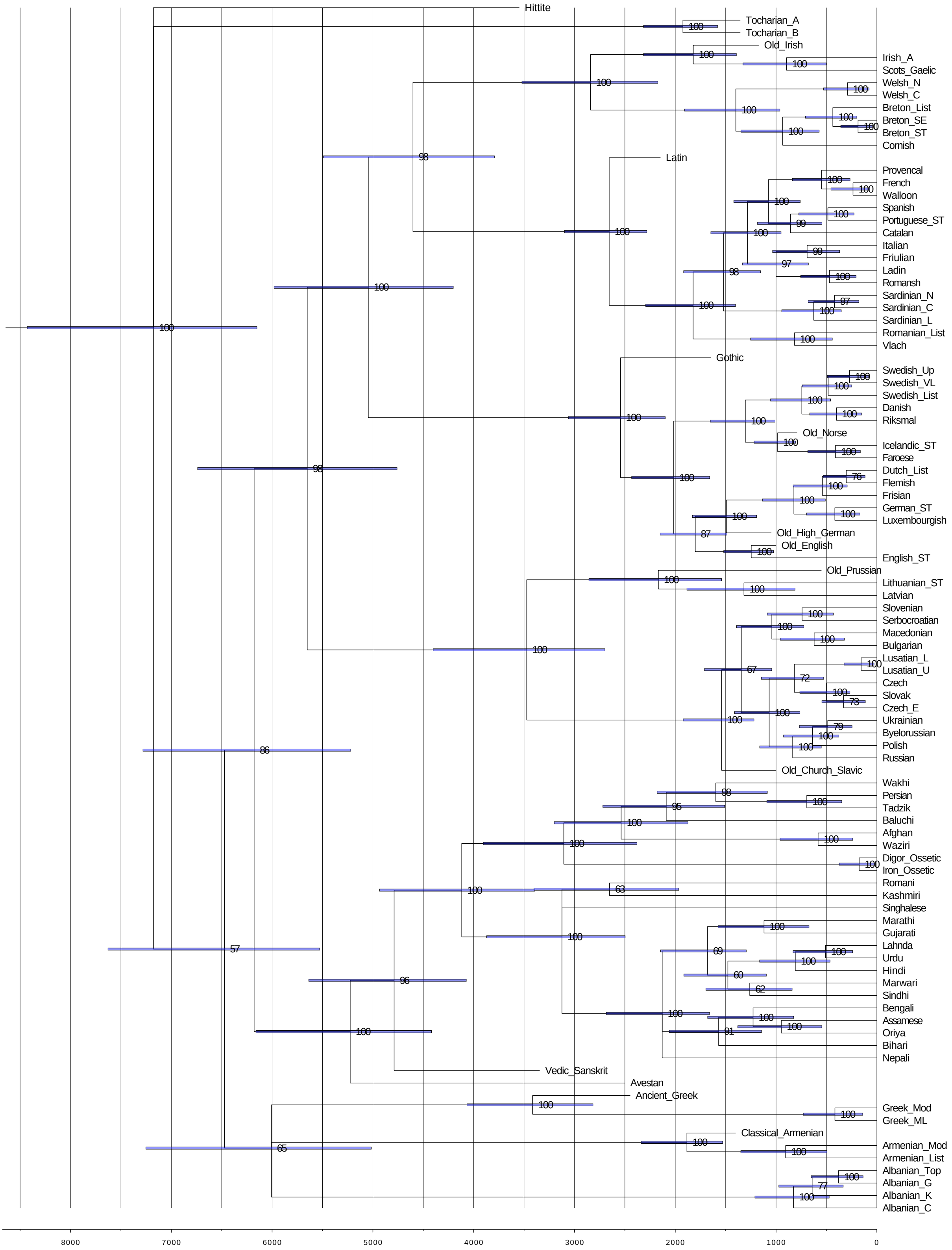


Figure 9: The majority-rule consensus tree for the B2 dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

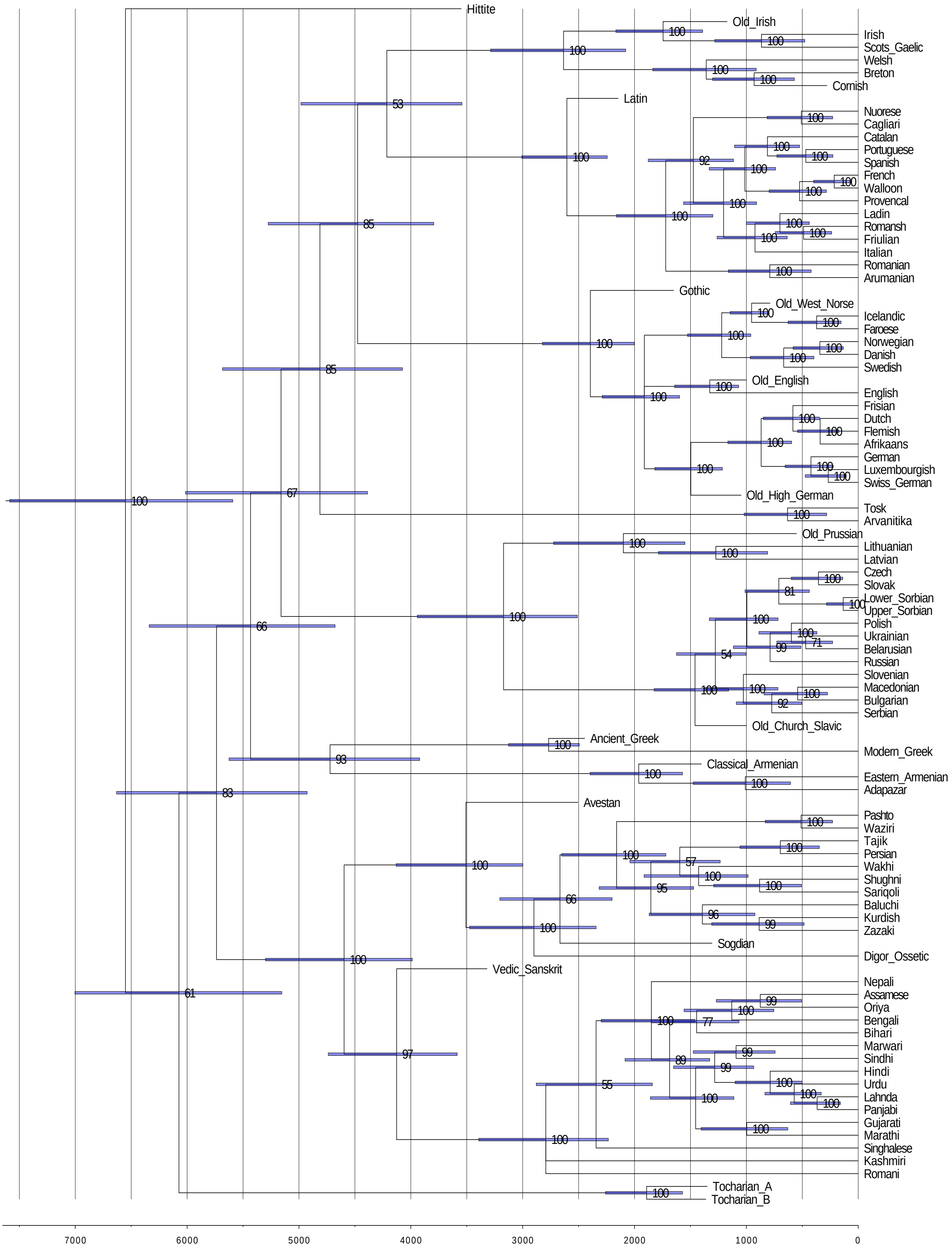


Figure 10: The majority-rule consensus tree for the BROAD dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

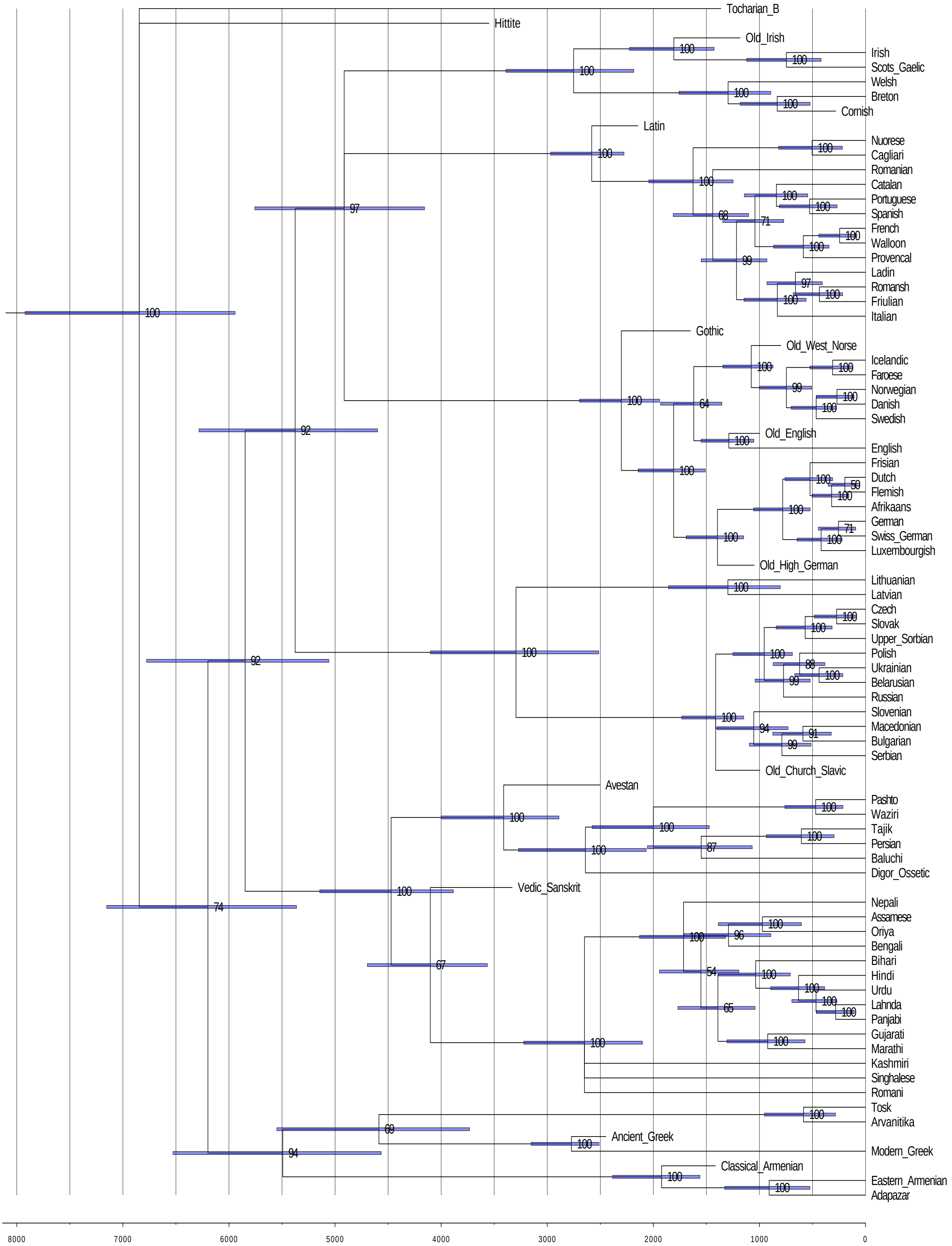


Figure 11: The majority-rule consensus tree for the MEDIUM dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

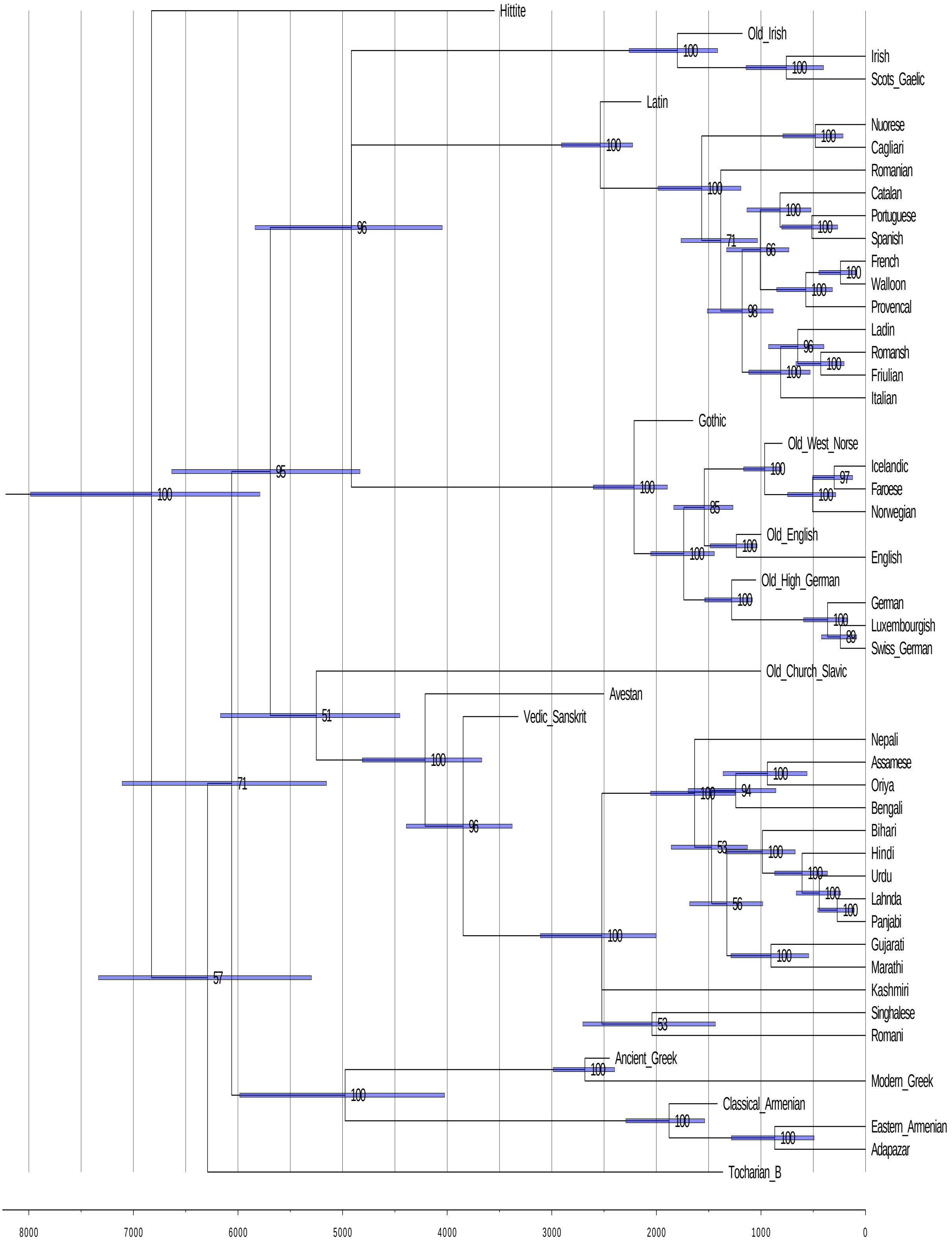


Figure 12: The majority-rule consensus tree for the NARROW dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

C Uniform Prior

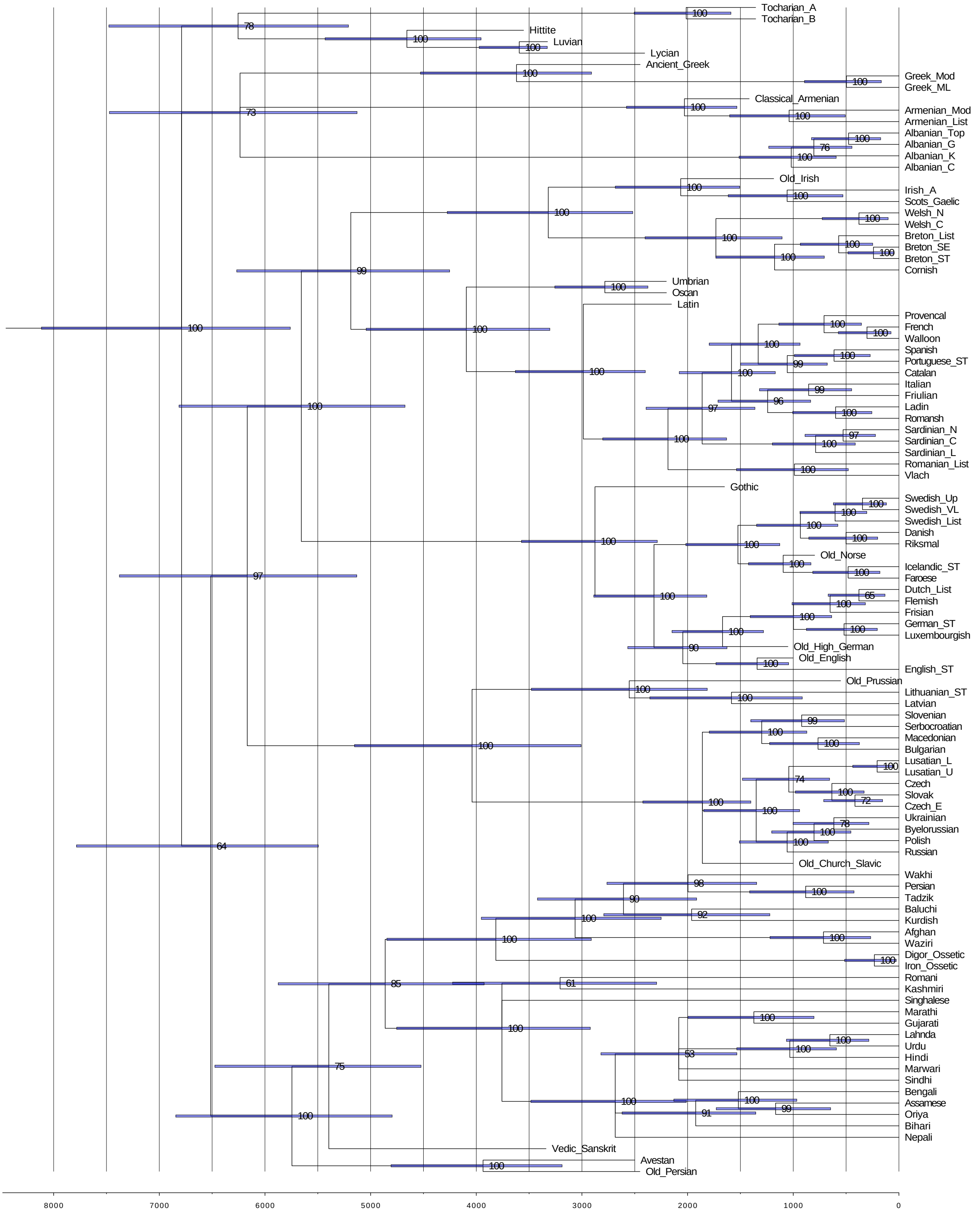


Figure 13: The majority-rule consensus tree for the B1 dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

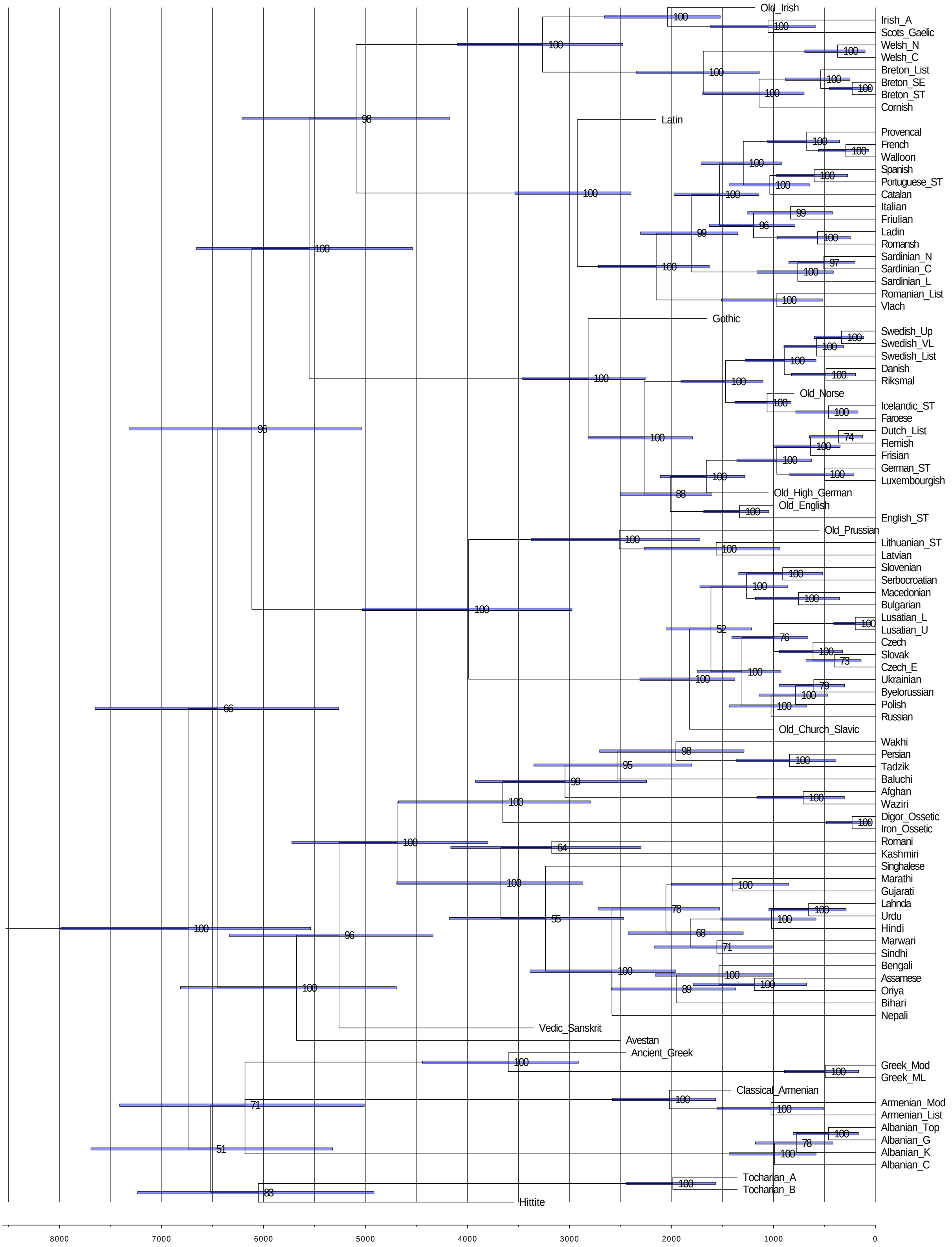


Figure 14: The majority-rule consensus tree for B2 dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

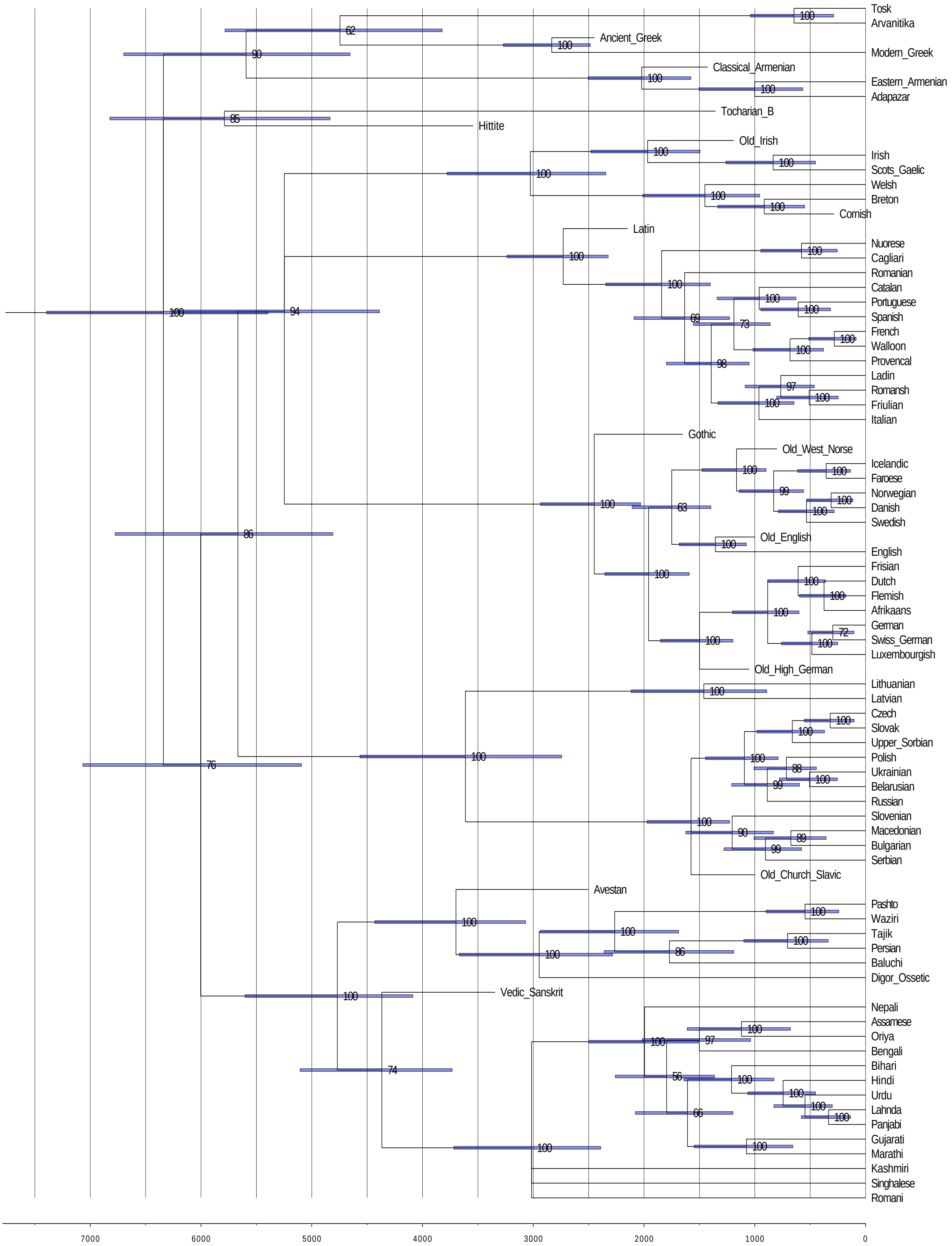


Figure 15: The majority-rule consensus tree for the MEDIUM dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.

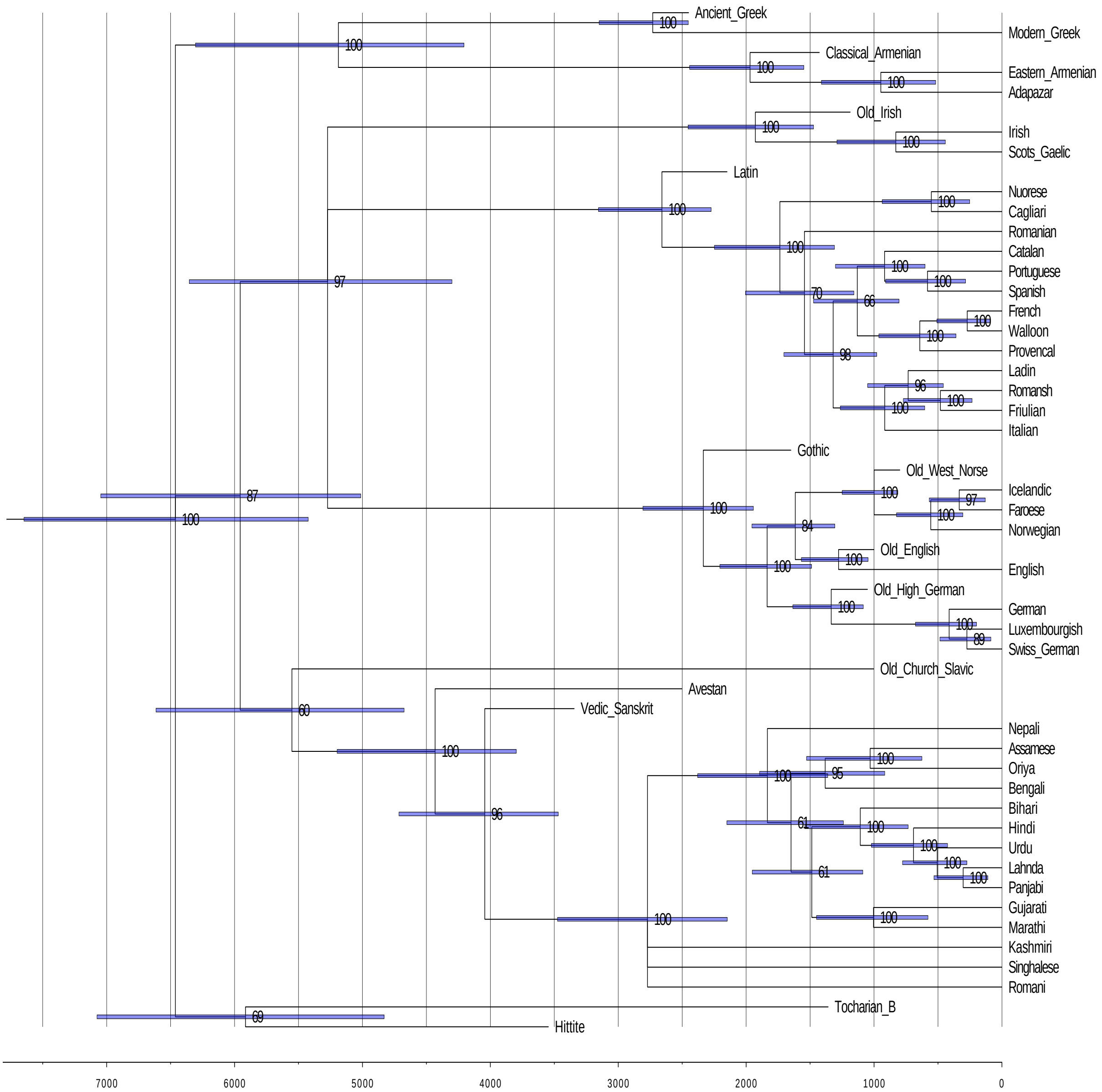


Figure 16: The majority-rule consensus tree for the NARROW dataset. The numbers at each internal node shows the support for the subtree in the posterior sample. The blue bars show the 95% HPD intervals for the node ages. The time scale shows the height of the tree in terms of age.