

ANCESTRY-CONSTRAINED PHYLOGENETIC ANALYSIS SUPPORTS THE INDO-EUROPEAN STEPPE HYPOTHESIS

WILL CHANG

University of California, Berkeley

CHUNDRRA CATHCART

University of California, Berkeley

DAVID HALL

University of California, Berkeley

ANDREW GARRETT

University of California, Berkeley

Discussion of Indo-European origins and dispersal focuses on two hypotheses. Qualitative evidence from reconstructed vocabulary and correlations with archaeological data suggest that Indo-European languages originated in the Pontic-Caspian steppe and spread together with cultural innovations associated with pastoralism, beginning c. 6500–5500 BP. An alternative hypothesis, according to which Indo-European languages spread with the diffusion of farming from Anatolia, beginning c. 9500–8000 BP, is supported by statistical phylogenetic and phylogeographic analyses of lexical traits. The time and place of the Indo-European ancestor language therefore remain disputed. Here we present a phylogenetic analysis in which ancestry constraints permit more accurate inference of rates of change, based on observed changes between ancient or medieval languages and their modern descendants, and we show that the result strongly supports the steppe hypothesis. Positing ancestry constraints also reveals that homoplasy is common in lexical traits, contrary to the assumptions of previous work. We show that lexical traits undergo recurrent evolution due to recurring patterns of semantic and morphological change.*

Keywords: lexical change, linguistic phylogenetics, Indo-European chronology, Indo-European dispersal, steppe hypothesis

This article has three main goals. First, we show that statistical phylogenetic analysis supports the traditional steppe hypothesis about the origins and dispersal of the Indo-European language family. We explain why other similar analyses, some of them widely publicized, reached a different result. Second, for skeptics about phylogenetic methodology, we suggest that the agreement between our findings and the independent results of other lines of research confirms the reliability of statistical inference of reconstructed chronologies. Finally, for linguistic phylogenetic research, we argue that analyses grounded in the evolutionary properties of the traits under study yield more reliable results. Our discussion makes reference to ancestry relationships, for example between Old Irish and two modern languages descended from it, Irish and Scots Gaelic, and draws on what can be learned from direct observation of changes over historical time. In our phylogenetic analyses, we introduce ANCESTRY CONSTRAINTS and show that they result in more realistic inferences of chronology.

Our article is organized as follows. We first give background information about the steppe and Anatolian hypotheses, and about earlier phylogenetic analyses (§1), and discuss lexical traits (§2) and linguistic ancestry relationships (§3). We then describe our data and some measurements made directly on the data (§4), explain our phylogenetic methods (§5), and summarize our experimental results (§6). Finally, we discuss the ef-

* We thank Claire Bowern, Russell Gray, Simon Greenhill, Hannah Haynie, Gary Holland, Lev Michael, Johanna Nichols, Donald Ringe, and Michael Weiss for comments and discussion, Tom Recht for help with Greek data, Quentin Atkinson and Remco Bouckaert for sharing code and data, Michael Dunn for assistance with IELEX, and audiences in Berkeley, Oslo, and San Diego for feedback and suggestions. Our work was supported by the Diebold Fund for Indo-European Studies, University of California, Berkeley (CC, WC, AG) and a Google Ph.D. Fellowship and Grant IIS-1018733 from the National Science Foundation (DH).

fects of advergence (§7) and ancestry constraints in phylogenetic modeling (§8), followed by conclusions (§9) and appendices with details about methods and results.¹

1. INDO-EUROPEAN BACKGROUND.

1.1. THE STEPPE AND ANATOLIAN HYPOTHESES. The relationships of Indo-European (IE) languages have been studied for over two centuries, but it is still disputed when and where their common ancestor Proto-Indo-European (PIE) was spoken, and how they spread before they first appeared in historical records about 3,700 years ago. Two hypotheses dominate discussion.

According to a traditional hypothesis (Gimbutas 1973, 1997, Mallory 1989) accepted by many linguists (Ringe 2006, Parpola 2008, Fortson 2010, Beekes 2011), PIE was spoken in the Pontic-Caspian steppe, north of the Black and Caspian Seas. The steppe hypothesis associates IE language spread with the diffusion of cultural innovations relating to pastoralism, including horse domestication, wheeled vehicles, and the weaving of wool from woolly sheep. Analyses of archaeological data from this point of view suggest a PIE dispersal date c. 6500–5500 BP, probably in the first half of that period.² It is now also widely assumed that Anatolian was the first branch to separate from PIE.³ Within the framework of the steppe hypothesis the common ancestor of the non-Anatolian languages, Proto-Nuclear-Indo-European (PNIE), might then have been spoken c. 6000–5000 BP.

According to an alternative hypothesis proposed by Renfrew (1987), IE languages spread into Europe with the diffusion of agriculture from Anatolia; see also Renfrew 1999, 2000a,b, 2001, 2003. This mechanism is plausible, since clear cases of language dispersal with the spread of agriculture are known elsewhere in the world (Bellwood & Renfrew 2002, Diamond & Bellwood 2003, Bellwood 2004). Given that farming reached southeast Europe by the seventh millennium BCE (van Andel & Runnels 1995, Perlès 2001, Bocquet-Appel et al. 2009), the Anatolian hypothesis implies a PIE dispersal date c. 9500–8000 BP.⁴

¹ Data sets and figures showing summary trees for all of our analyses appear in online supplementary materials, which can be accessed at <http://muse.jhu.edu/journals/language/v091/91.1.chang01.html>. We cite PIE verb roots from Rix et al. 2001, and we use the following abbreviations: BP: before present (2000 CE), HPD: highest posterior density, IA: Indo-Aryan, IE: Indo-European, IELEX: Indo-European Lexical Cognacy Database (<http://ielex.mpi.nl/>), ME: Middle English, OE: Old English, PIE: Proto-IE, NIE: Nuclear IE (the non-Anatolian IE clade), PNIE: Proto-NIE, RM: root-meaning (traits), RSC: restriction site character, SDC: stochastic Dollo character, (T)MRCA: (time of the) most recent common ancestor.

² Archaeologists have suggested PIE dates ‘about 4,500 BC’ (Mallory 1989) and ‘about 4,400–4,200 BC’ (Anthony 2013); see also Darden 2001, Mallory & Adams 2006, Anthony 2007, and Anthony & Ringe 2015. Some phylogenetic analyses attribute PIE dates of 6000–5000 BP to the steppe hypothesis (Gray & Atkinson 2003, Atkinson et al. 2005, Bouckaert et al. 2012); Ryder and Nicholls (2011) characterize the hypothesis more accurately.

³ This idea has a long pedigree under the label INDO-HITTITE (Sturtevant 1929, 1933) and is accepted in one form or another in much current research (Oettinger 1986, Strunk 1994, Melchert 1998, Lehrman 2001, Melchert 2001, Ringe et al. 2002, Jasanoff 2003, Rieken 2009, Yakubovich 2010).

⁴ A third hypothesis that has gained less traction can be seen as a hybrid of the Anatolian and steppe hypotheses. It posits that PIE was spoken at the time assumed by the steppe hypothesis, but in eastern Anatolia (Barber 2001) or the Caucasus (Gamkrelidze & Ivanov 1995, Ivanov 2001). According to this hypothesis, after PNIE spread north from Anatolia, the steppe was the staging ground for NIE expansion into Europe and Asia. All else being equal, however, linguistic geography favors a western origin for the Anatolian branch of IE, since the greatest branch-internal diversity is in the west; this is most easily explained by assuming that Proto-Anatolian spread across Anatolia from the west. Ivanov’s (2001) alternative analysis of Anatolian dialectology has not gained support from other specialists (Melchert 2001, Yakubovich 2010).

In principle, evidence bearing on IE origins and dispersal may come from archaeology, genetics, or linguistics. At present, genetic data is insufficient to resolve the matter, since ancient European DNA and comparison of ancient and modern DNA confirm not just immigration from the Near East at the time of the farming dispersal, but also later population movement from northern Eurasia that is consistent with the steppe hypothesis; see Brandt et al. 2013 and Lazaridis et al. 2014. In practice, discussion has mainly focused on archaeological and linguistic arguments.

Several arguments have been advanced in favor of the steppe hypothesis; a recent review is by Anthony and Ringe (2015). The first argument is from archaeological analysis. For example, based on correlations among archaeological data, documented cultural practices, and vocabulary, researchers have argued that Proto-Indo-Iranian was spoken c. 4300–3700 BP in an area of central Asia around the Aral Sea (Lubotsky 2001, Witzel 2003, Kuz'mina 2007).⁵ This includes the Sintashta culture of the steppe to the north, whose economy was pastoral and whose cemeteries contain horse sacrifices and chariots (Anthony 2009), as well as the more urbanized Bactria-Margiana Archaeological Complex to the south (Hiebert 1994). The latter may have been the staging ground for Indo-Iranian dispersal. If this argument is correct, then from what we can infer about cultural interactions in this region, Indo-Iranian speakers probably entered the area from the steppe. This line of reasoning locates speakers of an Indo-Iranian precursor north of the Caspian Sea c. 5000–4500 BP, close in time and place to PNIE if the latter was spoken in the steppe c. 6000–5000 BP. In other words, the diffusion of cultural traits that are observed in the archaeological record (and in some cases reported in later textual sources) correlates well with the chronology of the steppe hypothesis. Similar analyses link IE-speaking European populations with culture changes that can be identified as moving from the steppe and eastern Europe within the chronological framework of the steppe hypothesis.

A second argument is based on inferences about environment and material culture from reconstructed vocabulary. For example, wheeled-transport vocabulary is reconstructed for PIE (Mallory & Adams 1997, 2006, Parpola 2008) or PNIE (Darden 2001). Since wheeled transport was invented long after farming reached Europe, if PIE or PNIE had such vocabulary it cannot have been spoken by early farmers. It is fair to say that most of those writing from a linguistic perspective, though not all (Krell 1998, Clackson 2000), have been impressed by the extent of the evidence. The arguments are in any case based on an assemblage of individual points, each of which needs careful evaluation.

A third argument is based on linguists' subjective impression that early IE languages are more similar grammatically and phonologically than would be expected from the Anatolian chronology; see Table 1. After 4,500 or more years of divergence on the Anatolian chronology, some grammatical patterns remain intact with only a few changes in each language; see, for example, Hittite [e:smi, e:si, e:stsi] = Sanskrit [ásmi, ási, ásti]. Similarly, only a few sound changes distinguish Hittite [χanti], Sanskrit [ánti], and Greek [ánti]; such examples can be replicated throughout the grammar and lexicon. The

⁵ Because language and material culture spread independently, some archaeologists emphasize that correlations of specific archaeological data with ethnicity or language cannot be demonstrated conclusively without inscriptional evidence. But most linguists would probably agree that Kohl (2009:236) goes too far, in his otherwise excellent survey of Bronze Age Eurasia, in suggesting that '[t]here was no single Indo-European or Proto-Indo-European "homeland" but just an ever unfolding historical process of development'. If there was a PIE language, it was spoken somewhere.

	PIE	HITTITE	VEDIC SANSKRIT	GREEK
'I am'	*h ₁ ésmi [ʔésmi]	ēšmi [e:smi]	ásmi [ásmi]	eimí [e:mi]
'you (sg.) are'	*h ₁ ési [ʔési]	ēšši [e:si]	ási [ási]	eí [ē:] (dialectal <i>essí</i>)
's/he is'	*h ₁ ésti [ʔésti]	ēšzi [e:stsi]	ásti [ásti]	estí
'they are'	*h ₁ sénti [ʔsénti]	ašanzi [asantsi]	sánti [sánti]	eisí [e:si] (dialectal <i>entí</i>)
'bear'	*h ₂ rtkós [ʔrtkos]	ḫartaggaš [ʔartkas]	ḫkṣas [ḫkṣas]	árktos
'cloud, sky'	*néb ^h os [néḫos]	nēpiš [ne:bis]	náb ^h as [náḫas]	nép ^h os
'wood'	*dóru [dóru]	tāru [ta:ru]	dāru [dā:ru]	dóru
'vis-à-vis'	*h ₂ ánti [ʔánti]	ḫanti [ʔanti]	ánti [ánti]	antí
'yoke'	*yugóm [jugóm]	yugan [jugan]	yugám [jugám]	z _u gón

TABLE 1. Selected vocabulary in PIE and three early IE languages.

divergence time posited by the Anatolian hypothesis is roughly twice that of the present-day Germanic, Romance, or Slavic languages, but many linguists have a subjective impression that the differences in Table 1 are not twice as great. Yet we have no generally accepted way to quantify impressions of similarity in phonology or grammar, or to show that examples like those in Table 1 are representative, so this argument remains impressionistic.⁶

Two further arguments probably support only a weaker position, namely, that PNIE was spoken in the Pontic-Caspian steppe. One concerns evidence for early contact between IE and early western Uralic languages (Joki 1973, Koivulehto 2001, Janhunen 2009). Given the location of Uralic in northern Eurasia, such contact must have occurred north of the Black and Caspian Seas. The evidence for contact between early Uralic languages and the Indo-Iranian branch of IE is uncontroversial, supported by dozens of unambiguous loanwords, and accepted by specialists (Rédei 1986, Lubotsky 2001, Mallory 2002). There is similarly clear evidence for contact with Balto-Slavic (Kallio 2005, 2006, 2008) and Germanic (Hahmo et al. 1991–2012). The evidence for contact with PIE itself is weaker (Kallio 2009), perhaps because Uralic languages spread from the east into northern Europe and Proto-Uralic itself was not spoken in proximity to the steppe.⁷

Finally, some morphological evidence suggests that the Greek, Armenian, Balto-Slavic, and Indo-Iranian subfamilies form a clade within IE (Ringe et al. 2002). Since Greek is spoken to the west of Anatolia, and Armenian and Indo-Iranian to the east, it is hard to construct a diversification scenario consistent with the Anatolian hypothesis in which these languages remained in contact after PNIE, unless the latter was itself spoken on the steppe. The steppe hypothesis makes this easier: the four subfamilies in question remained in proximity, after the departure of Tocharian to the east and Italo-Celtic, Germanic, and others to the west.

Two main arguments support the Anatolian hypothesis. First, as originally noted by Renfrew (1987), the spread of agriculture provides a plausible mechanism for large-scale language dispersal, one with clear parallels elsewhere. The language dispersal mechanisms required in the steppe hypothesis are less well understood, partly because pastoral subsistence economies are not as common worldwide. Second, beginning with Gray & Atkinson 2003, the Anatolian hypothesis has been supported by research using

⁶ Holman and colleagues (2011) provide a phonological distance measure from which they infer a PIE time-depth of about 4350 BP. That is far too late (only a few centuries before the first attestation of already differentiated IE languages), but it does quantify the impression that IE languages are more similar phonologically than in vocabulary patterns. See Clackson 2000 for a perceptive, skeptical discussion of the argument from impressionistic similarity.

⁷ For recent discussion see Häkkinen 2012 and Parpola 2012.

statistical methods adapted from biological phylogenetics (Atkinson et al. 2005, Atkinson & Gray 2006, Nicholls & Gray 2008, Gray et al. 2011, Ryder & Nicholls 2011, Bouckaert et al. 2012, 2013). This is also the focus of our research.

Methodological differences between fields contribute to the present impasse. Arguments for the steppe hypothesis are mostly qualitative rather than quantitative, and come from traditional lines of reasoning in historical linguistics and archaeology. In contrast, a crucial argument for the Anatolian hypothesis is quantitative, relying on statistical methods that originated in another discipline. Thus some researchers have expressed skepticism about chronological inference with statistical methods (Clackson 2000, 2007, Evans et al. 2006, McMahon & McMahon 2006), while some advocates of such methods have written that historical linguistics methods ‘all ... involve intuition, guesswork, and arguments from authority’ (Wheeler & Whiteley 2014). We hope our work can contribute to a rapprochement between the two research traditions.

1.2. INDO-EUROPEAN PHYLOGENETICS. IE linguistic phylogeny has been studied for many decades (Meillet 1922, Porzig 1954, Birnbaum & Puhvel 1966), but statistical phylogenetic research is relatively recent in IE (Tischler 1973). Dyen and colleagues (1992) used lexicostatistics to produce a classification of IE languages by analyzing a word list of eighty-four modern languages and 200 basic meanings compiled by Isidore Dyen. Their method assumed a similar overall rate of lexical change in all languages. An alternative approach to classification that dispenses with this assumption was employed by Ringe and colleagues (2002) to address the issue of higher-order structure in IE. They analyzed a data set created by Don Ringe and Ann Taylor, consisting of phonological, morphological, and lexical traits from twenty-four predominantly ancient and medieval languages. These two works yielded two word lists with cognate coding; one or both, or both combined, were used in all subsequent work.

In 2003, Gray and Atkinson presented the first Bayesian phylogenetic analysis of IE chronology; they used the Dyen word list, supplemented with Hittite, Tocharian A, and Tocharian B data. Historically attested events were used to date various linguistic splits, and RATE SMOOTHING over the branches of the inferred tree was used to relax the assumption of a constant rate of change. As in all subsequent analyses prior to our work, the inferred root age supports the Anatolian hypothesis. Nicholls and Gray (2008) reworked Gray and Atkinson’s analysis by replacing the trait model, which permitted multiple gains in the same lexical trait, with one that did not; they also performed a separate analysis on the lexical traits of a subset of the languages from the Ringe-Taylor data set. Ryder and Nicholls (2011) then added a model of lexicographic coverage that enabled them to work with all twenty-four languages in the Ringe-Taylor data set, many of which, like Oscan and Old Persian, are scantily attested. Bouckaert and colleagues (2012) performed a phylogeographic analysis: its goal was to infer the geographical location of PIE, but embedded in it was a phylogenetic analysis that superseded previous work in most respects.⁸ Most notably, the inference software supported many different trait models, including the single-gain trait model devised by Nicholls and Gray, and the data was based on a harmonization of the Ringe-Taylor and Dyen data sets. Bouckaert and colleagues (2013) addressed an error in the coding of the data without altering their general conclusions.

⁸ Our work to date does not model geographic dispersal or address their phylogeographic findings; this remains a desideratum for future research.

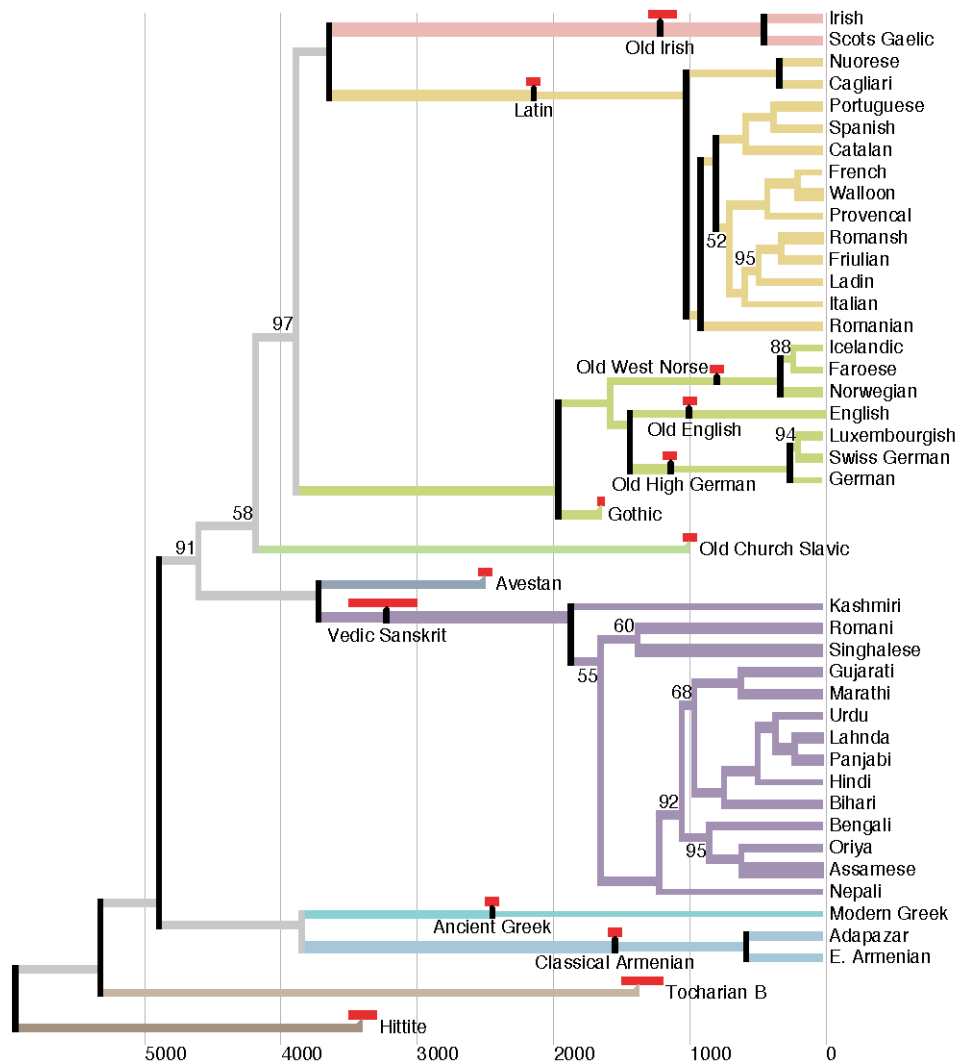


FIGURE 1. Analysis A1 summary tree. Modern languages with no ancestors in the data set are excluded. This tree shows median posterior node heights, median posterior branch rate multipliers (width of horizontal lines), time constraints on ancient and medieval languages (bright red bars), clade constraints (vertical black bars), and posterior clade probabilities less than 98%.

We now briefly preview our results. Using the same model and data set as Bouckaert and colleagues (2012, 2013), but with incremental changes to both, we found a root age that strongly supports the steppe hypothesis. The key difference was that we constrained eight ancient and medieval languages to be ancestral to thirty-nine modern descendants. Using ancestry constraints is similar in spirit to stipulating the known dates of historical languages, or stipulating uncontroversial clades that are not the object of inquiry. The ancestor-descendant relationships we posit are uncontroversial, but could not be inferred by the model. Figure 1 shows the result of an analysis with ancestry constraints and other refinements, where the only modern languages included are those with documented ancestors. Figure 2 shows a similar analysis with modern languages

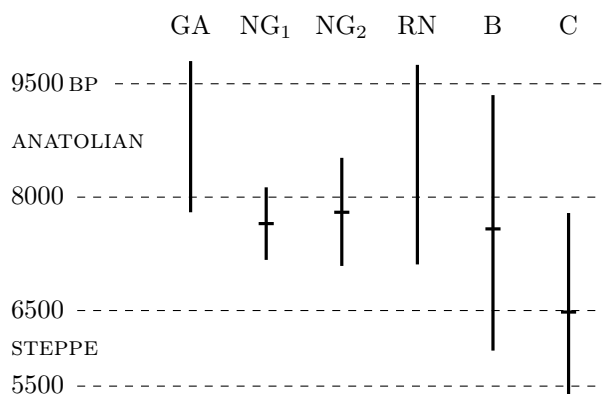


FIGURE 3. Inferred IE root age distributions in selected studies. GA: Gray & Atkinson 2003; NG₁, NG₂: Nicholls & Gray 2008, using Dyen and Ringe-Taylor data sets; RN: Ryder & Nicholls 2011; B: Bouckaert et al. 2013; C: analysis A1 corrected root age from our work (§7.1). Plotted are the 95% highest-density interval (vertical lines), the mean (NG, RN) or median (B, C) if known, and intervals for the steppe and Anatolian hypotheses (dashed lines).

in ways that are poorly understood.⁹ For this reason, and because large lexical data sets are available, most statistical work on language relationships analyzes lexical traits. There are at least two types of lexical traits, which we call COGNATE TRAITS and ROOT-MEANING TRAITS. They have not been distinguished in previous phylogenetic research.

Languages share a cognate trait if they share cognate words, that is, words descended from the same ancestral word form. For example, English and German share a cognate trait because *timber* and *Zimmer* ‘room’ are descended from Germanic **timra-* (derived from a PIE root **demh₂-* ‘build’); likewise, German *Gast* ‘guest’ and Latin *hostis* ‘stranger, enemy’ define a cognate trait because they are descended from a form **g^hosti-* (Bammesberger 1990, Ringe 2006). Cognate words need not have the same meaning. Cognate traits are widely studied in comparative and historical linguistics, but are only occasionally used in statistical phylogenetic studies (e.g. Gray & Jordan 2000).¹⁰

More often, the data consists of root-meaning (RM) traits, which encode whether the most semantically general and stylistically neutral word for a given meaning is based on a given ancestral root; meanings are often chosen from a ‘Swadesh’ list of one or two hundred basic meanings. Such traits are the basis for most IE analyses, including ours. For example, since English *feather* is derived from a PIE root **pet-* ‘to fly’, English has a trait [**pet-*, ‘feather’]; because Latin *serpens* ‘snake’ is derived from **serp-* ‘to creep’,

⁹ Typological traits are also sometimes used in analyzing language relationships, though they share drawbacks of both lexical traits (they spread easily) and nonlexical traits (they are often interdependent). Some difficulties of using morphological and phonological traits in statistical phylogenetic analyses have also been discussed by Taylor and colleagues (2000) and Ringe and colleagues (2002).

¹⁰ Cognate traits can be hard to use systematically, since they require detailed etymological knowledge. In English, for example, without medieval data a cognate trait would be missed by anyone who did not know that the second member of the compound *bridegroom* is the cognate of Gothic *guma* ‘man’; the cognate of Sanskrit *mādhu-* ‘honey’ would be missed without lexical data that included the infrequent word *mead*. A further problem in the analysis of cognate traits is to determine which words count as ‘descended from the same ancestral word form’. In Romance, for example, Latin neuters change gender, and simplex verbs are often continued by their frequentatives. Presumably Italian *rapa* ‘turnip’ counts as the ‘same’ as Latin *rāpum*, though *rapa* strictly speaking continues a plural *rāpa* reinterpreted as singular; but is French *chanter* < Latin *cantāre* ‘sing (repeatedly)’ the ‘same’ as Latin *canere* ‘sing’? Statistical analyses of root-meaning traits need not seek in vain for a principled way to untie such Gordian knots.

Latin has a trait [**serp-*, ‘snake’]. Languages can share an RM trait based on forms with different derivations, like English *feather* < **pet-trā* (or **pet-rā*) and Latin *penna* ‘feather’ < **pet-nā*. These words share a root but are not cognate words, since they were derived with unrelated suffixes **-trā* (or **-rā*) and **-nā* and so cannot descend directly from the same ancestral word form. By contrast, because the cognate words *timber* and *Zimmer* have different meanings, they do not define a shared RM trait.

2.1. HOMOPLASY AND DRIFT. Cognate and RM traits evolve very differently, especially with respect to HOMOPLASY or independent innovation.¹¹ Except in borrowing between languages, cognate traits ordinarily come into existence only once; this is the basis of the COMPARATIVE METHOD (Meillet 1925, Weiss 2014). Therefore models of trait evolution that do not permit homoplasy are well suited to cognate traits. But because the mechanisms of change underlying RM traits include semantic change and the derivation of new words from existing forms, RM traits are subject to at least two distinctive kinds of homoplasy. In describing them, we adapt Sapir’s (1921) term DRIFT, which refers to the predisposition to undergo certain changes given certain precursor traits.

First, RM traits arise not only when word forms come into existence, but also when they change meanings. For example, Old English (OE) *timber* probably originally meant ‘building’ (like Old Saxon *timbar*); to model its shift in meaning to ‘timber (material for building)’, the trait [**demh₂₋*, ‘building’] would be said to be replaced by a trait [**demh₂₋*, ‘timber’].¹² Meaning changes fall into recurrent patterns across languages (Heine & Kuteva 2002, Traugott & Dasher 2002, Urban 2014). If the same meaning change affects the same root in related languages, a homoplastic RM trait results. For example, in a crosslinguistically common shift (Wilkins 1996), reflexes of PIE **pod-* ‘foot’ came to mean ‘leg’ independently in Modern Greek and modern Indic and Iranian languages. Two other examples from our data set are given in 1.

- (1) a. Old Irish *seinnid* meant ‘play or strike an instrument, sound’ but has shifted in Modern Irish and Scots Gaelic to mean ‘sing’. The ancestral root **swenh₂₋* referred to producing sound or music more generally, but the same semantic shift to ‘sing’ is seen in Persian *x^vāndan*.
- b. Many languages distinguish a stative ‘sit’ verb (‘be in a sitting position’) from a change-of-state one (‘sit oneself down, take a seated position’), but shifts between the senses are common. In PNIE, the root **h₁eh₁s-* expressed the stative sense, while **sed-* expressed the change-of-state sense (Rix et al. 2001). Change-of-state **sed-* came to express the stative sense independently in Armenian, Balto-Slavic, Celtic, Germanic, and Italic, a shift not shared by Greek or Indo-Iranian and therefore independent in at least some of the branches where it happened.

Recurrent meaning changes like these have been called ‘rampant’ in language (Ringe et al. 2002). For such changes we use the term SEMANTIC DRIFT.

¹¹ Homoplasy is an evolutionary term for independent analogous innovation in parallel lineages. Changes like the *t > k* sound change that has occurred independently at least twenty times in Austronesian languages (Blust 2004) are said to be homoplastic.

¹² We write as if later meanings simply replace earlier ones. In reality, as in this case (Old English had both meanings), words usually pass through a polysemous stage, which would be modeled as the coexistence of traits (e.g. [**demh₂₋*, ‘building’] and [**demh₂₋*, ‘timber’]). The methodology of basic-vocabulary word list collection tends to suppress all but the clearest cases of polysemy in the languages under study, but this should not affect the analysis as long as the method is the same throughout the data set.

A second source of homoplasy in RM traits is DERIVATIONAL DRIFT. This refers to change that occurs because certain roots are semantically well suited to provide certain derivatives. For example, constructions that mean ‘cause to die’ are a recurrent source of verbs for ‘kill’ (Buck 1949). Therefore, as descendants of PIE **g^{wh}en-* ‘kill’ fell out of use, causative derivatives of PIE **mer-* ‘die’ were used in this meaning. This happened independently in Irish and in modern Indic and Iranian languages (Rix et al. 2001), yielding a homoplastic RM trait.¹³ Three other examples from our data set are given in 2.

- (2) a. Because words for ‘animal’ often evolve from expressions meaning ‘having breath’ and so forth, PIE *h₂enh₁-* ‘breathe’ is seen in several otherwise unrelated ‘animal’ terms. Latin *animal* itself is a derivative of *anima* ‘spirit’, a derivative of *h₂enh₁-*. In Indo-Iranian, Persian *jānvār* ‘animal’ and related forms descend from **wyāna-bāra-* ‘having a spirit’, in which **wyāna-* is a derivative of *h₂enh₁-*. And though not the basic word for ‘animal’, Tocharian B *onolme* ‘living being’ is also a derivative of *h₂enh₁-* ‘breathe’ (D. Adams 2013).
- b. Basic words for ‘live’ include derivatives of the PNIE root **g^wyeh₃-*: Greek *zōō*, Classical Armenian *keam*, Latin *vīvō*, and so forth, all of which are primary verbal formations. Basic Celtic words for ‘live’ in our data set are derivatives of an adjective **g^wih₃-wo-* that was derived from the same root, that is, a construction ‘be alive’. It is natural to derive a stative verb ‘X’ from a construction ‘be X’ (with a stative adjective X).
- c. In the meaning ‘snake’, reflexes of the PIE noun **h₁óg^{wh}is* are widespread (Ancient Greek *óphis*, Vedic Sanskrit *áhi-*, etc.). A verb **serp-* ‘crawl’ also often refers specifically to the motion of a snake. Derivatives of **serp-* came to be the general term for ‘snake’ in Albanian, in Latin (and modern Romance languages), and in modern Indo-Aryan (IA) languages, for example, Hindi *sāp*, Assamese *xāp*. The homoplastic nature of such cases is shown by the fact that the various word forms that acquire the new meaning are often formed with different derivations. For example, though Albanian (Tosk) *gjarpër* ‘snake’ and Latin *serpens* are both based on the PIE root **serp-*, the Albanian noun is formed with a suffix **-ena-* (Orel 1998) and the Latin noun with a different suffix **-ent-*. The word forms themselves do not go back to a single ancestor.

In short, though some analysts erroneously assume that RM traits are homoplasy-free (Atkinson et al. 2005:204, Gray et al. 2011:1094), semantic and derivational drift are endemic in RM data sets. This claim is further supported in §3 below and quantified in §4.2.¹⁴

¹³ Derivational drift can be understood as resulting from the semantic drift of derivative words. For example, the derivational drift whereby ‘cause to die’ replaces ‘kill’ can be viewed as a consequence of semantic change in these verbs. It is nonetheless helpful to distinguish the two categories when possible.

¹⁴ A third kind of lexical trait, intermediate between cognate and RM traits, would encode whether words for given meanings are expressed by cognate word forms, rather than roots. Such traits would have evolutionary properties like those of RM traits, since they would be liable to semantic drift (though not derivational drift). These are evidently the traits analyzed by Bowerman and Atkinson (2012), who report that ‘words were not counted as cognate ... if they were present in the language in a different meaning’ (p. 827). They also write that ‘languages are highly unlikely to independently gain the same cognate’ (p. 829), which is true but irrelevant, since (as they implicitly acknowledge by noting that cognates can have different meanings) the traits they analyze are susceptible to semantic drift.

2.2. PRECURSOR TRAITS AND ADVERGENCE. A precondition for semantic and derivational drift is the presence of PRECURSOR traits: those that tend to give rise to semantic changes or new derivatives as described in §2.1. These are also heritable RM traits. For example, a language must have a precursor trait like [**pod-*, ‘foot’] in order to innovate [**pod-*, ‘leg’] via semantic drift; it must have a precursor like [**serp-*, ‘crawl’] to innovate [**serp-*, ‘snake’] via derivational drift. In traditional terms, Ancient Greek *poús* must have meant ‘foot’ (not ‘mouth’ or ‘stone’) in order to gain the meaning ‘leg’, and Latin *serpō* must have had a suitable meaning for a derivative to mean ‘snake’. Some precursor traits, like [**pod-*, ‘foot’], are in the basic-vocabulary data set, but most, like [**serp-*, ‘crawl’], are not. Basic words do not make a closed semantic system: the semantic sources of basic vocabulary are often elsewhere.

A precursor trait may cause a modern descendant of an ancestral language to gain a corollary trait in parallel with a nondescendant. For example, Latin *homō* ‘person’ and Old Irish *duine* ‘person’ are both derived from PIE **dʰǵʰom-* ‘earth’ (Ernout & Meillet 1951, Matasović 2009); the derivation is like that of *earthling*. Thus Latin and Old Irish share a trait [**dʰǵʰom-*, ‘person’].¹⁵ Semantic shift in both Romance and Modern Irish produced a corollary trait [**dʰǵʰom-*, ‘man’]: reflexes of *homō* shifted from ‘person’ to ‘man (male person)’, as in French *homme*, and a parallel change gave Modern Irish *duine* ‘man’.

Interacting with precursor-driven drift is a further phenomenon, which Renfrew (2000a) called ADVERGENCE: ‘the process of mutual influence when two separate languages, which are in fact genetically related through descent from a common ancestor, occupy adjacent territories and continue to interact’. In other words, diversifying languages that remain in contact in a dialect network tend to share innovations (Geraghty 1983, Ross 1988, Babel et al. 2013). Three examples in our data set are given in 3. In 3b–c, note that the West Scandinavian language Norwegian has had intensive and sustained contact with East Scandinavian.¹⁶

- (3) a. Latin *edere* ‘eat’ was replaced in that meaning by reflexes of *mandūcāre* ‘chew’ (e.g. French *manger*) in all Romance languages outside of Iberia. Since these languages are contiguous but paraphyletic, contact must have played a role in causing them to gain the trait [**mand-*, ‘eat’], but this also occurred because Latin had a precursor trait [**mand-*, ‘chew’]: words for ‘chew’ are a regular source of words for ‘eat’ (Buck 1949), and in colloquial speech *mandūcāre* was occasionally used to mean ‘eat’ (Glare 1982).
- b. Germanic languages generally have the cognate of English *eat* in the sense ‘eat’, including Old West Norse *eta*, preserved in Icelandic; but Danish and Norwegian use *spise*.
- c. For the sense ‘narrow’ the Old West Norse word was *mjór* ‘thin, narrow’ or *pröngr* ‘narrow, crowded’, but Danish, Norwegian, and Swedish have *smal*.

¹⁵ Since the formations are morphologically distinct (Celtic **gdonyo-* with suffix *-yo-* vs. Italic **dʰǵʰemon-* with suffix *-on-*), derivational drift is a possible alternative account. Similar formations occur not only in the Sabellic branch of Italic (e.g. Oscan *humuns* ‘people’), but also in Germanic (e.g. OE *guma* ‘man, person’) and Balto-Slavic (e.g. Old Lithuanian *žmuō*); Fortson (2010:25) reconstructs the trait for PIE.

¹⁶ Dutch and Low German also played a significant role in this process, better described as advergence among North Sea Germanic languages. In both examples in 3b–c, Low German was the source of the innovated words.

We use Renfrew's term in a broader sense. When closely related languages accumulate the same innovations, not present in their common ancestor, it can be hard to determine how much of this is due to drift (because the languages share precursor traits) and how much is also caused by sociolinguistic interaction. We call all such developments advergent here.

2.3. BORROWING. While our analyses (like many phylogenetic studies) use lexical data, the full picture of a language family's phylogeny is also based on other data. In fact, lexical data is usually judged less reliable than other linguistic data as an indicator of phylogeny (Campbell & Poser 2008, Ringe & Eska 2013). Any linguistic feature can be borrowed given sufficiently intense contact (Thomason 2001), but vocabulary is most readily borrowed. Loanwords are thus often excluded in analyses whose goal is to determine relationship, including those that seek to infer phylogeny and chronology together (Gray & Atkinson 2003, Bouckaert et al. 2012, 2013). But in analyzing chronology alone, loanword exclusion is not motivated. First, the replacement of one word by another in the same semantic slot is the same kind of change whether the new item is a native word or was borrowed at some previous time. Loanwords do not normally enter a language in the basic vocabulary; they may enter as stylistic alternants, only later settling into their eventual semantic slot. For example, the French borrowing *animal* took several hundred years to become the most general and neutral English word in the meaning 'animal' (Kurath & Kuhn 1952–2001).

Moreover, known borrowings are not evenly distributed across IE languages. For modern languages like English and French and medieval languages like Old Irish and Old Norse, we know most languages they have been in contact with and we can recognize borrowings. But for ancient languages like Vedic Sanskrit and Ancient Greek, given our poor knowledge of their linguistic milieu it is almost impossible to distinguish borrowings from distributionally limited inheritances. For example, the Ancient Greek word for 'woods' is *húlē*. This word has many derivatives within Greek but no secure cognates elsewhere in IE, nor any obvious Greek morphological source (Chantraine 1999, Beekes 2010). It might have been borrowed from a language that Greek came in contact with, or it might be an internal creation whose history is obscure; we cannot know. Such cases also arise at the subfamily level. For example, the basic Germanic word for 'drink' (English *drink*, Old Norse *drekka*, etc.) has no IE cognates and no known origin. The uneven distribution of known loanwords is quantified in §4.2.

If known loanwords are excluded from the data, then the measured rate of change, which is inferred largely from modern and medieval languages, will be lower (cf. Greenhill et al. 2009). Since unidentified ancient loanwords cannot be excluded, the unattenuated diversity in ancient languages, combined with the lower inferred rate of change, will cause the root of the tree to be placed farther in the past. Thus it is more principled not to exclude tagged loanwords, even though in practice the effect of doing so is quite small (see Fig. 5 below, analysis A4).

3. ANCESTRY RELATIONSHIPS. Table 2 lists the eight ancient and medieval languages in our data set that are known to be the ancestors of thirty-nine modern languages in the data set, and for which we implemented ancestry constraints.

A logically possible alternative is that the later languages are descended not from the putative ancestor, but from another variety spoken at the same time. For example, perhaps Modern Irish and Scots Gaelic are descended not from Old Irish, but from an undocumented variety that had already significantly diverged from it.

This alternative is the only interpretation of the phylogenetic trees given by Bouckaert and colleagues (2012, 2013), whose analyses include ancestral and descendant languages

ANCESTOR	DESCENDANT LANGUAGE OR CLADE (AND SOURCES)
Vedic Sanskrit	IA languages (Turner 1962–1966, Masica 1991, Cardona & Jain 2003)
Ancient Greek	Modern Greek (Meillet 1920, Horrocks 1997)
Latin	Romance languages (Väänänen 1981, Herman 2000, J. N. Adams 2007, 2013, Clackson & Horrocks 2011, Wright 2011)
Classical Armenian	Modern Armenian dialects: Adapazar, Eastern Armenian (Hübschmann 1897, Adjarian 1909, Vaux 2006)
Old Irish	Goidelic: Irish, Scots Gaelic (Jackson 1951, Ahlqvist 1988)
Old English	English (Luick 1921–1940)
Old West Norse	West Scandinavian: Faroese, Icelandic, Norwegian (Bergsland & Vogt 1962, Haugen 1976)
Old High German	High German: German, Swiss German, Luxembourgish (Salmons 2012)

TABLE 2. Ancestry relationships in our data. Sources summarize the relevant descent relationships.

but do not constrain their relationship except that each ancestral language forms a clade with its descendants. In each such clade their results show extensive lexical change on the branch leading from the common ancestor to the ancestral language, which is then crucially not ancestral to the later descendants. For example, their maximum clade credibility tree shows Old Irish evolving for over 500 years after its common ancestor with the other Goidelic languages (for a similar effect see our Fig. 6 below).

The phylogeny implied by Bouckaert and colleagues' results would mean either that the putative ancestral languages in Table 2 were highly diglossic, with undocumented colloquial varieties possessing many basic-vocabulary differences from documented varieties, or that similar basic-vocabulary differences characterized undocumented regional dialects of the languages in Table 2. As far as we know, these claims have not featured in the scholarly literature; there are several good arguments against them, which we summarize in this section.

First, where we know about regional dialects of the ancient and medieval languages in Table 2, there is no evidence for significant basic-vocabulary differences. For example, the Ancient Greek dialects were diverse and numerous and are well documented throughout the Greek-speaking world, since the written language was not standardized until the Hellenistic period. But even thorough descriptions of Ancient Greek dialect variation (Buck 1955, Colvin 2007) do not mention relevant differences in basic vocabulary; new lexical traits in Modern Greek are the result of historical change over two millennia. Admittedly, we do not know about the regional dialects of every language in Table 2, but it would be tendentious to propose the needed highly divergent dialects only in those cases where we are ignorant.

Second, where we know about social variation in the ancient and medieval languages in Table 2, there is no evidence for significant basic-vocabulary differences. The best-understood case is Latin, where evidence of the colloquial language is found in private letters, contracts, graffiti, and other documents, as well as literary depictions and grammarians' comments (Väänänen 1981, Herman 2000, J. N. Adams 2007, 2013, Wright 2011). Colloquial speech in any language has distinctive properties, but the evidence does not show that literary and colloquial Latin would differ in a Swadesh list. As Clackson and Horrocks (2011:236) describe Latin in the first three centuries CE, 'there was so much geographical and social mobility' that 'local differences in speech tended to become levelled, and long-term divisions in the language were kept to a minimum'.¹⁷

¹⁷ Wright (2011:63) emphasizes that the familiar term *VULGAR LATIN* is 'a collective label for those features of Latin which we can be sure did exist, but which were not recommended by the grammarians. It should not be, although it often has been, envisaged as being a separate language (or "system") coexisting with "Classical" Latin'.

To confirm the implausibility of a Romance ancestor contemporary with Latin but lexically quite divergent from it, we examined all cases in our data set where a Latin word was replaced in its meaning category by another word in at least twelve of the fourteen Romance languages we analyzed. On the hypothesis of extreme diglossia, these would be cases where colloquial Latin speech was already lexically distinct. There are thirteen examples. Three are in grammatical (specifically deictic) vocabulary: Latin *hic* ‘this’, *ille* ‘that’, and *ibi* ‘there’ were replaced by grammaticalized phrasal expressions in Romance languages. All three words cited are documented in Latin sources of all styles and time periods; even in colloquial Latin, the vocabulary replacements had not happened (Lodge 1924–1933, Väänänen 1981, Glare 1982). The remaining ten examples are in content vocabulary. As seen in Table 3, in each case the record is clear that while the semantic change that eventually led to replacement was sometimes nascent in Latin, even in colloquial sources the original word was the most general and stylistically neutral word in the relevant meaning category. Literary and colloquial Latin had the same basic vocabulary.

ALL	<i>L omnis</i> ‘all’ was replaced by reflexes of <i>tōtus</i> ‘every, the entirety of’. But <i>tōtus</i> is not semantically general in L: it usually modifies singulars and rarely means ‘(they) all’, while <i>omnis</i> modifies plural nouns and is common in all registers.
FAT	<i>L adeps</i> ‘fat, grease’ was replaced by reflexes of <i>crassum</i> ‘thick, stout’. But <i>adeps</i> is used in plain style, including agricultural treatises; <i>crassum</i> (the source of Italian <i>grasso</i> ‘fat, grease’, etc.) had extended meanings (of liquids) ‘concentrated, muddy, thick’ and (metaphorically of people) ‘insensitive, crass’. The semantic shift from ‘fat (of people)’ to ‘fat, grease’ is natural and crosslinguistically common, as in English.
FIRE	<i>L ignis</i> ‘fire’ was replaced by reflexes of <i>focus</i> ‘hearth, fireplace’. But <i>ignis</i> was used in all registers and styles, including colloquial sources (e.g. Plautus); numerous derivatives also refer to fire. By contrast, <i>focus</i> was used in the sense ‘hearth, fireplace’ in all registers and styles, including colloquial sources (e.g. Plautus). The metonymic shift ‘fireplace’ > ‘fire’, first observed in late technical writings (Väänänen 1981:98), was natural when <i>ignis</i> fell out of usage in the post-Roman period.
MAN	<i>L vir</i> ‘man, adult male’ was replaced by reflexes of <i>homō</i> ‘person’. The former is very well documented in all registers and styles, including colloquial sources (e.g. Plautus, Petronius, graffiti), whereas <i>homō</i> means ‘person’ and not ‘adult male’, including in colloquial sources. When <i>vir</i> was lost in late L and early Romance, <i>homō</i> shifted to fill its semantic position.
MOUTH	<i>L ōs</i> ‘mouth’ was replaced by reflexes of <i>bucca</i> ‘jaw, lower part of the cheeks’. But <i>ōs</i> is very well documented in all registers and styles, including authors with colloquial vocabulary (e.g. Plautus, Cato, Terence). By contrast, <i>bucca</i> means ‘jaw, lower part of the cheeks’ even in colloquial sources and is only sometimes used to mean ‘mouth’; Väänänen (1981:78) calls it a ‘familiar and expressive synonym’ of <i>ōs</i> , that is, not stylistically neutral. The JAW > MOUTH semantic shift is of a crosslinguistically common metonymic type (Buck 1949).
PERSON	<i>L homō</i> ‘human being, person’ was replaced by reflexes of <i>persōna</i> ‘a mask, a dramatic character, a person’s character, persona, or personality, legal person’. The older word <i>homō</i> is used in the meaning ‘person’ in all stylistic registers, including colloquial sources (e.g. Plautus, Terence, Petronius), while the Etruscan loanword <i>persōna</i> has a restricted set of specialized meanings in all L registers.
SKIN	<i>L cutis</i> was replaced by reflexes of <i>pellis</i> ‘(animal) hide, treated skin’. The common semantic shift ‘hide, (animal) skin’ > ‘(human) skin’ (Buck 1949) occurred early in Romance and is reflected in all of the modern Romance languages (e.g. French <i>peau</i> , Italian <i>pelle</i>). The word <i>pellis</i> when used in Latin sources (including colloquial ones) to refer to human skin implied ‘an unkempt or uncultivated condition’.
SWELL	<i>L tumēre</i> ‘to swell’ was replaced by derivatives of <i>flāre</i> ‘to blow’. But <i>tumēre</i> was used in all registers and styles, including by authors with colloquial language (e.g. Plautus); it has many derivatives expressing the same sense. Many Romance languages use derivatives of <i>flāre</i> ‘to blow’ in the sense ‘swell’. But <i>L flāre</i> never meant ‘swell’, and the Romance languages do not agree on which derivative of <i>flāre</i> they use. For example, Italian <i>gonfiare</i> is from <i>L conflāre</i> ‘ignite, stir up’, while Catalan <i>inflar</i> is from Latin <i>inflāre</i> ‘blow on, inflate, puff up, cause to swell’; lack of agreement means that the developments were separate.

(TABLE 3. *Continues*)

WARM	<i>L tepidus</i> ‘warm (not hot)’ was replaced by reflexes of <i>calidus</i> ‘hot’; <i>tepidus</i> has many derivatives in the sense ‘warm’, while <i>calidus</i> means ‘hot’ in all registers and styles.
WIDE	<i>L lātus</i> ‘broad, wide’ was replaced by reflexes of <i>L largus</i> ‘munificent, generous’. But <i>lātus</i> means ‘wide’ even in colloquial authors (e.g. Plautus) and epigraphic sources, and the same meaning is absent or extremely rare for <i>largus</i> in any source.

TABLE 3. Romance advergence. Shown are all ten content-word meaning categories in our data set where a Latin (L) word was replaced in its meaning category by another word in at least twelve of the fourteen Romance languages we analyzed. Data from Väänänen (1981:75–84) and Glare (1982).

Two other arguments concern homoplastic traits, that is, traits shared by languages in more than one IE subfamily but not by a putative ancestor in one subfamily. In some cases we have textual data showing how word meanings shift. In such cases it is possible to show that the homoplastic traits arise independently through internal processes or borrowing. For example, English has a homoplastic trait in the word *belly*, since its Irish cognate *bolg* also means ‘belly’ but the OE word for ‘belly’ was *wamb*. The Bouckaert et al. phylogeny might explain this as an inheritance from the common ancestor of Celtic and Germanic, transmitted through an unattested OE sibling. But text evidence shows that OE *belg* meant only ‘bag, bellows’ (Cameron et al. 2007) and its ME successor *belī* ‘bellows’ came to mean ‘stomach’ (by 1225) and then ‘abdomen, belly’ (by 1395; Kurath & Kuhn 1952–2001). This pattern of metaphorical and then metonymic extension is semantically natural, so there is no reason to attribute the ‘belly’ sense to an otherwise undocumented variety of English. Similarly, English has a homoplastic trait in the word *fat* ‘(animal) fat, grease’ (cf. German *Fett*). From an original adjectival use ‘fattened, fat’ in both Old and Middle English (Cameron et al. 2007, Kurath & Kuhn 1952–2001), this has come to refer to animal fat only in Modern English through a metonymic shift (perhaps promoted by advergence involving other Germanic languages). There is no reason to posit an undocumented OE variety in which this word already had the modern use.

Finally, even where we may lack detailed text evidence for the key stages, the homoplastic traits in question have explanations involving well-established patterns such as derivational and semantic drift. For example, ignoring borrowed words, our data set includes twenty-five traits found in modern IA languages and in non-IA languages, but not in Vedic Sanskrit (the language of the Vedas, and the most archaic attested form of Sanskrit). If Sanskrit were not the ancestor of IA languages, we could assume that these traits were inherited from an earlier IE stage via an undocumented Vedic Sanskrit sibling. Certainly Vedic had unrecorded variation, inferred from later evidence, but this has crucially not been identified for basic-vocabulary traits.¹⁸ We examined all modern IA homoplastic traits in our data set, setting aside known borrowings. All twenty-five examples reflect derivational or semantic drift.¹⁹ If they were inherited from an unattested Vedic Sanskrit sibling, there would be no reason to expect Sanskrit to have a pre-

¹⁸ On inferred variation in the Vedic period see Emeneau 1966, Hock 1991, Norman 1995, Oberlies 1999, 2001, and Witzel 1999. We should emphasize that we do not mean to imply that Vedic Sanskrit (or any language) lacked a full repertoire of stylistic variants and partial synonyms. But the object of study here is quite limited: it is the set of the most semantically general and stylistically neutral words for a set of basic meanings; there is no evidence for what we below call OVERLOADED basic-vocabulary traits.

¹⁹ Sometimes their distribution points to advergence; Indo-Aryan is a well-known area of contact-induced spread. Masica (1991:60) writes that Dardic, Romani, and Singhalese are the only IA languages or groups that ‘were beyond the pale of the evolving pan-Indic civilization and thus unaffected by M[iddle] IA changes that spread to all contiguous languages participating in the latter’.

cursor for each trait. But in fact, each homoplastic trait has a plausible precursor in Sanskrit, and each historical change from precursor to corollary trait is plausible.

Eight traits in IA languages are homoplastic due to derivational drift. A good example involves the meaning ‘kill’ (also mentioned in §2.1). In this meaning the basic Vedic Sanskrit verb was *han-* < PIE **ǵ^{wh}en-* ‘kill’. But in several modern IA languages (e.g. Hindi *mārnā*) the basic verb for ‘kill’ is derived from the PIE root **mer-* ‘vanish, die’, as in some non-IA languages (e.g. Scots Gaelic *marbh*). The IA verbs are derived from a causative form of that root (‘cause to die’), which is not reconstructed for PIE (Rix et al. 2001) but did exist in Vedic Sanskrit; it meant ‘cause to die’ and was not the basic word for ‘kill’.²⁰ The change is a natural one that is liable to occur independently. The remaining seven examples of IA homoplasmy due to derivational drift are given in 4: in 4a is a verb derived from a noun; the other examples are nouns and adjectives derived from verbs, usually denoting prototypical agents, instruments, or undergoers.²¹

(4) Indo-Aryan homoplastic traits: Derivational drift

- a. TOOTH → BITE (Vedic *dánt-* ‘tooth’ → Romani W *dandel*; cf. Persian *dandān giriftan*, etc.; the basic Vedic verb for ‘bite’ was *damś-*)
- b. BE HOT → WARM (Vedic *tap-* ‘be hot’ → Romani W *tato*, etc.; cf. Avestan *tapta-*; the basic Vedic word for ‘warm’ was *uṣṇá-*)
- c. CRAWL → SNAKE (Vedic *sarp-* ‘crawl, creep’ → Hindi *sāp*, etc.; cf. Latin *serpēns*, etc.; the basic Vedic word for ‘snake’ was *áhi-*)
- d. DECAY → OLD (Vedic *jar-* ‘decay, wither’ → Gujarati *jūnū*, etc.; cf. Modern Greek *géros*, etc.; the basic Vedic word for ‘old’ was *sána-*)
- e. DIE → MAN (Vedic *mar-* ‘die’ → Romani W *murš*; cf. Persian *mard*, etc.; the basic Vedic word for ‘man’ was *nár-*); or a combination of derivational drift DIE → PERSON in 4f below and semantic drift PERSON > MAN in 5c.iv below
- f. DIE → PERSON (Vedic *mar-* ‘die’ → Kashmiri *murth*; cf. Avestan *mašīia-*, etc.; the basic Vedic word for ‘person’ was *púruṣa-*)
- g. SEE → EYE (Vedic *kaś-/caḥ-* ‘see’ → Bengali *cok(h)*; cf. Persian *čašm*, etc.; though infrequent in the Ṛg Veda, the basic Vedic word for ‘eye’ was *ákṣi-*²²)

An additional seventeen traits in IA languages are homoplastic due to semantic drift. A good example involves the meaning ‘leg’ (also mentioned in §2.1). In this meaning the basic Vedic Sanskrit noun was *jānghā-*, but several modern IA languages have a

²⁰ This is shown both by frequency (it is much rarer than *han-*) and by examples like this one from Atharva Veda (Śaunaka recension) 13.3.3a, where it is explicitly coordinated with a second causative verb: *yó māráyati prāṇáyati* ‘[the deity Rohita], who causes [beings] to die, and to breathe’. Note that Ossetic *mar-* must continue a Proto-Iranian **mār-aya-* (cf. *mæl-* ‘die’ < **m₁-ya-*); it is formally possible for this and the IA forms to be cognate. Causatives are very productive in Indo-Iranian, though, so independent derivational drift is also plausible. Crucially, the Indo-Iranian forms are only indirectly related to Scots Gaelic *marbh* (derived from Celtic **marwā-* ‘die’ (Matasović 2009)) and Latin *murenter* ‘kill’ (derived from Latin *morī* ‘die’ with one or more suffixes, reflecting zero-grade **m₁-*). Differences in ablaut and morphology show that these verbs do not warrant positing an archaic trait [**mer-*, ‘kill’] and instead have come about independently via derivational drift.

²¹ Cited in each case in 4–5 are representative IA and non-IA examples of the same innovated trait, usually one of each. IA forms are quoted from Grierson 1932, Turner 1962–1966, or ROMLEX (Halwachs et al. 2007), with Romani varieties abbreviated as follows: U = Ursari, W = Welsh.

²² The Vedic term *cakṣus-* is more frequent overall, but has a broad range of meanings (e.g. ‘seeing, shining, sun’), is less common in the meaning ‘eye’ (Grassmann 1873), and was not (yet) the basic word in that meaning.

word for ‘leg’ (e.g. Bengali *pā*) whose PIE ancestor is reconstructed with the meaning ‘foot’ because that is the only meaning any reflex has in any ancient IE language. Outside of IA, the meaning ‘leg’ is found in modern languages like Greek (*pōdi*) and Persian (*pā*), but Sanskrit *pād-* means only ‘foot’. The trait [**pod-*, ‘leg’] shared by several modern IE languages represents a crosslinguistically well-documented metonymic shift ‘foot’ > ‘leg’ (Wilkins 1996). In the Middle IA language Pāḷi, *janghā* means ‘leg’, and *pada* still exclusively means ‘foot’ (Davids & Stede 1921–1925, Turner 1962–1966). This shows that this semantic change postdates the Middle IA period, and that the trait [**pod-*, ‘leg’] is not an archaism. The remaining sixteen examples of IA homoplasy due to semantic drift are given in 5. Where there is clear evidence, a Middle IA form (from Pāḷi or Prakrit) is given to confirm that the meaning attested in Vedic was preserved in a later period, and that the semantic innovations seen in one or more modern IA languages had not yet taken place.

(5) Indo-Aryan homoplastic traits: Semantic drift

- a. Aktionsart shifts: Change-of-state and activity verbs become stative.
 - i. TREMBLE > FEAR (Vedic *tras-* ‘tremble’ > Romani W *trašel*; cf. Persian *tarsīdan*, etc.; the basic Vedic verb for ‘fear’ was *bhay-*, still continued by Pāḷi *bhay-* ‘fear’ vs. *tasati* ‘tremble, fear’)
 - ii. PERCEIVE > KNOW (Vedic *jñā-* ‘perceive, recognize’ > Hindi *jānnā*, etc.; cf. Ossetic (Digor) *zon-*, etc.; the basic Vedic verb for ‘know’ was *ved-*)
 - iii. FALL ASLEEP > BE SLEEPING (Vedic *svap-* ‘fall asleep’ > Hindi *sonā*, etc.; cf. Avestan *xvafsa-*, etc.; the basic Vedic verb for ‘be sleeping’ was *sas-*)
- b. Metonymic and metaphorical shifts
 - i. HIDE > BARK (Vedic *cārman-* ‘hide’ > Kashmiri *tsam*; cf. Ossetic (Digor) *c’aræ*; the basic Vedic word for ‘bark’ was *tvác-*, still continued by Pāḷi *tacō* ‘bark, skin’ vs. *camma* ‘leather, shield’)
 - ii. HIDE > SKIN (Vedic *cārman-* ‘hide’ > Hindi *cām*, etc.; cf. Sogdian *crm*, etc.; the basic Vedic word for ‘bark’ was *tvác-*)
 - iii. CLOUD > SKY (Vedic *nábhas-* ‘cloud’ > Kashmiri *nab*; cf. Old Church Slavonic *nebo*, etc.; the basic Vedic word for ‘sky’ is *dyáus-*)
 - iv. BATHE > SWIM (Vedic *snā-* ‘bathe’ > Romani *najol*; cf. Latin *nāre*, etc.; the basic Vedic verb for ‘swim’ was *plav-*, still continued by Prakrit *pavai* ‘swims’ vs. *ṇhāi* ‘bathes’)
 - v. FOREST > TREE (Vedic *vána-* ‘forest’ > Sindhi *vaṇu*; cf. Avestan *vana*, etc.; the basic Vedic word for ‘tree’ was *vṛkṣá-*, still continued by Prakrit *vakkha* ‘tree’ vs. *vaṇa* ‘forest’)
 - vi. HEAT > SUN (Ved *gharmá-* ‘heat, warmth’ > Romani W *kham*; cf. Old Irish *grían*, etc.; the basic Vedic word for ‘sun’ was *súvar-*)
 - vii. CORD > ROPE (Vedic *raśmí-* ‘cord’ > Panjabi *rassī*, etc. < Sanskrit *raśmí*; cf. Zazaki *resen*, with a different suffix, equivalent to Sanskrit *raśanā-* (Mayrhofer 1989–2001); the basic Vedic verb for ‘rope’ was *rájju-*)
- c. Semantic generalization (i–iii) and specialization (iv–vi)
 - i. BLOW > BREATHE (Old Indic *phūt-kr-* ‘blow’ > Romani W *phurdel*, etc.; cf. Classical Armenian *p’c’em*; the basic Vedic verb for ‘breathe’ was *an-*)
 - ii. ANCIENT > OLD (Vedic *purāṇa-* ‘ancient’ > Hindi *purānā*, etc.; cf. Old High German *firni*, etc.; the basic Vedic word for ‘old’ was *sána-*)

- iii. CRUSH > RUB (Old Indic **mar-* ‘crush, rub’ (Turner 1962–1966) > Nepali *malnu*, etc.; cf. Old Irish *con-meil*, etc.; the basic Vedic verb for ‘rub’ was *gharṣ-*)
- iv. PERSON > MAN (Vedic *mānu-* ‘person, (Primordial) Man’ > Assamese *mānuh*, etc.; cf. English *man*, etc.; the basic Vedic word for ‘man’ was *nár-*)
- v. CUT > SPLIT (Vedic *ched-* ‘cut (off)’ > Romani U *čhinel*; cf. Latin *scindit*, etc.; the basic Vedic verb for ‘split’ was *bhed-*, still continued by Pāṇi *bhindati* ‘splits, breaks’ vs. *chēdana* ‘cutting’)
- vi. GO > WALK (Vedic *yā-* ‘go’ > Marwari *jā*; cf. Hittite *iyatta*, etc.; the basic Vedic verb for ‘walk’ was *car-*)

In short, the data supports our view that the ancient and medieval languages in Table 2 are the ancestors of their listed modern descendants. Contrary to the analysis of language relationships that is entailed by Bouckaert et al. 2012, 2013, it is not necessary or even plausible to assume undocumented colloquial or sibling language varieties with markedly distinct basic lexica.

4. DATA AND MEASUREMENTS.

4.1. FROM WORD LIST TO TRAIT MATRIX. We drew our data from IELEX, a database created and curated by Michael Dunn. This began as a harmonization of the Dyen and Ringe-Taylor data sets (§1.2) and has since been edited and expanded. We analyzed a selection of IELEX data that we extracted on April 21, 2013, and edited to correct Indo-Iranian coding errors. This data set is available in the online supplemental materials (<http://muse.jhu.edu/journals/language/v091/91.1.chang01.html>).

IELEX contains words for 207 meaning classes in over 150 IE languages. Within each meaning class, words are grouped by their IE root, so it is easy to construct RM traits from the data. In preparation for analysis, we assembled trait matrices in which each column encodes an RM trait. Each cell indicates the presence (1) or absence (0) of a particular trait in a particular language (Table 4). An ideal database would provide exactly one word for each SLOT (each pairing of meaning class and language), but often there are multiple or none. When IELEX lists multiple words in a slot, the slot is OVERLOADED, and we take each word to indicate a distinct trait (cf. Friulian in Table 4).²³ When IELEX provides no words for a slot, the slot is EMPTY, and we assume that the language in question could have any (none, one, or multiple) of the traits attested in the meaning class by any other language. The cells corresponding to an empty slot are marked HIDDEN (written ‘?’, as for Hittite in Table 4).

TRAIT	X	Y	Z	X	Y	Z
English	feather			1	0	0
Friulian	pene	plume		1	1	0
Spanish		pluma	→	0	1	0
Greek	ftero			1	0	0
Persian			par	0	0	1
Hittite				?	?	?

TABLE 4. Words from IELEX for ‘feather’ (left), coded as a trait matrix (right). The competing forms in Friulian result in an OVERLOADED SLOT. The lack of a form in Hittite yields an EMPTY SLOT; each corresponding cell, marked ‘?’, is a HIDDEN CELL.

In IELEX, recognized loanwords are tagged, so we can choose to exclude or include them from the trait matrix. In principle there are many ways to implement LOAN EX-

²³ Our analyses use binary traits, but some linguistic phylogenetic analyses use multistate traits, one per meaning, that code which root is found in the basic word. In this context, if multiple forms are given for the same slot, the trait is deemed POLYMORPHIC.

CLUSION; we follow Bouckaert et al. 2012 and put 0 in the cell of a tagged loanword (cf. German and Greek in Table 5, center). Note that this produces a different outcome from an empty slot (cf. Lycian). We implement LOAN INCLUSION by adhering literally to how RM traits are defined: if a language has a certain root in a certain meaning, then it has the trait, regardless of whether the word is a loan (Table 5, right).

TRAIT	X	Y		X	Y	X	Y
Spanish	fruta			1	0	1	0
German	(Frucht)			0	0	1	0
Greek	(fruto)		→	0	0	or 1	0
Russian		plod		0	1	0	1
Lycian				?	?	?	?

TABLE 5. Words from IELEX for ‘fruit’ (left), coded with loan exclusion (center) or loan inclusion (right). Loanwords are parenthesized.

	B	BROAD	MEDIUM	NARROW
Languages	103	94	82	52
Modern	83	78	69	39
Ancient and medieval	20	16	13	13
Meaning classes	207	197	143	143
Empty slots	1,745	601	45	33
Traits	5,992	5,694	3,279	2,350
Attestations	22,353	20,802	13,388	8,615
Hidden cells	49,346	20,433	1,049	484
Max in a language	4,649	3,504	70	48

TABLE 6. Four data sets. An EMPTY SLOT is when a language is unattested in a meaning class; an ATTESTATION is any cell in the trait matrix bearing 1; a HIDDEN CELL is any cell that corresponds to an empty slot; and B refers to Bouckaert et al. 2012, 2013.

NAME	DATES ASSUMED IN OUR ANALYSES
Hittite (hit)	1500–1300 BCE = 3400 ± 100 BP
Vedic Sanskrit (san)	1500–1000 BCE = 3250 ± 250 BP
Avestan (ave)	550–450 BCE = 2500 ± 50 BP
**Sogdian (sog)	600–800 CE = 1300 ± 100 BP
Ancient Greek (grc)	500–400 BCE = 2450 ± 50 BP
Latin (lat)	200–100 BCE = 2150 ± 50 BP
Gothic (got)	325–375 CE = 1650 ± 25 BP
Old High German (goh)	800–900 CE = 1150 ± 50 BP
Old English (ang)	950–1050 CE = 1000 ± 50 BP
Old West Norse (non, ‘Old Norse’)	1150–1250 CE = 800 ± 50 BP
Classical Armenian (xcl)	400–500 CE = 1550 ± 50 BP
**Tocharian A (xto)	500–800 CE = 1350 ± 150 BP
Tocharian B (txb)	500–800 CE = 1350 ± 150 BP
Old Irish (sga)	700–900 CE = 1200 ± 100 BP
*Cornish (cor)	1600–1800 CE = 300 ± 100 BP
Old Church Slavic (chu)	950–1050 CE = 1000 ± 50 BP
**Old Prussian (prg)	1400–1500 CE = 550 ± 50 BP

TABLE 7. Earlier languages in our analyses. Shown are language names we use, ISO 693-3 abbreviations, and in one case a language name used in IELEX. Cornish (*) is excluded from the narrow data set; languages with two asterisks (**) are excluded from the narrow and medium data sets; the broad data set includes all languages above (see text for explanation). Dates represent the earliest (or for Cornish latest) substantial attestation of basic vocabulary. Ancient Greek dates refer to the classical period, not earlier poetic dialects.

INDO-IRANIAN	<p>Indo-Aryan: Nepali (nep), Assamese (asm), Oriya (ori), Bengali (ben), *Bihari (mai), **Marwari (rwr), Hindi (hin), Urdu (urd), **Sindhi (snd), Lahnda (pnb), Panjabi (pan), Gujarati (guj), Marathi (mar), Kashmiri (kas), Singhalese (sin), Romani (Greek) (rmy, 'Gypsy Gk')</p> <p>Iranian: **Wakhi (wbl), *Baluchi (bgp), **Kurdish (kmr), **Zazaki (diq), *Tajik (tgk, 'Tadzik'), *Persian (pes), *Waziri (pst), *Pashto (pus), **Shughni (sgh), **Sariqoli (srh), *Ossetic (Digor) (oss, 'Digor Ossetic')</p>
ROMANCE	<p>Sardinian: Cagliari (sro, 'Sardinian C'), Nuorese (src, 'Sardinian N')</p> <p>Eastern Romance: Romanian (ron), **Arumanian (rup, 'Vlach')</p> <p>Italo-Western Romance: Catalan (cat), Portuguese (por, 'Portuguese St'), Spanish (spa), French (fra), Provençal (prv), Walloon (wln), Ladin (lld), Romansh (roh), Friulian (fur), Italian (ita)</p>
CELTIC	<p>British: *Breton (bre, 'Breton St'), *Welsh (cym, 'Welsh N')</p> <p>Goidelic: Irish (gle, 'Irish B'), Scots Gaelic (gla)</p>
GERMANIC	<p>North Germanic: Icelandic (isl), Faroese (fao), Norwegian (nob), *Swedish (swe), *Danish (dan)</p> <p>West Germanic: English (eng), *Frisian (frs), *Dutch (nld), *Flemish (vls), *Afrikaans (afr), German (deu), Luxembourgish (ltz, 'Lëtzebuergesch'), Swiss German (gsw, 'Schwyzerdütsch')</p>
BALTO-SLAVIC	<p>South Slavic: *Serbian (srp), *Bulgarian (bul), *Macedonian (mkd), *Slovenian (slv)</p> <p>East Slavic: *Russian (rus), *Belarusian (bel, 'Byelorussian'), *Ukrainian (ukr)</p> <p>West Slavic: *Polish (pol), *Slovak (slk), *Czech (ces), *Upper Sorbian (hsb), **Lower Sorbian (dsb)</p> <p>Baltic: *Latvian (lav), *Lithuanian (lit)</p>
OTHER	<p>Albanian: *Tosk (als, 'Albanian T'), *Arvanitika (aat, 'Albanian K')</p> <p>Armenian: Eastern (hye, 'Armenian Mod'), Adapazar (hye, 'Armenian List')</p> <p>Modern Greek: (ell, 'Greek Mod')</p>

TABLE 8. Contemporary languages in our analyses. Shown are language names we use, ISO 693-3 abbreviations, and in some cases language names used in IELEX. Languages with one asterisk (*) are excluded from the narrow data set; languages with two asterisks (**) are excluded from the narrow and medium data sets; the broad data set includes all languages above (see text for explanation). Because some IELEX word lists originated in the 1960s and some sources are from the mid-twentieth century or earlier, contemporary languages are constrained to 1950–2000 CE = 50–0 BP. The one exception is English, assigned a date of 2000 CE = 0 BP because some of the software that we use requires that at least one taxon be current.

From IELEX data we assembled three data sets (Table 6). The BROAD DATA SET consists of ninety-four languages and 197 meaning classes. Of the 207 meaning classes in IELEX, three ('blow', 'father', 'mother') were excluded as being especially susceptible to sound symbolism, and seven others were excluded because they were unattested in more than 30% of the languages, probably due to the fact that they are not in the Dyen data set, which was constructed around the 200 items proposed by Swadesh (1952). We also assembled two smaller data sets because of the possibility that the many sparsely attested languages in the broad data set could bias the analysis. Our MEDIUM DATA SET uses a subset of eighty-two languages and 143 meaning classes from the broad data set, chosen so that no language would have too many hidden cells (see the last row in Table 6). We assembled a NARROW DATA SET by removing from the medium data set any modern language that lacks an ancestor in the data set, leaving fifty-two languages. Finally, to follow up on the analyses of Bouckaert and colleagues (2012, 2013), we reconstructed the data set featured in their analyses, which they derived from IELEX at an earlier date. The languages we analyze are listed in Tables 7 and 8, and the clade constraints assumed in our analyses are listed in Table 9.

4.2. DATA SET MEASUREMENTS. Using our data sets, independent of phylogenetic analysis, we quantify claims made earlier: that homoplastic traits are common (§2.1),

Nuclear Indo-European: all languages other than Anatolian
Inner Indo-European: all languages other than Anatolian and Tocharian
Balto-Slavic: Baltic + Slavic
Baltic: Old Prussian + East Baltic
East Baltic: Latvian, Lithuanian
Slavic: East + South + West Slavic
West Slavic: Czech, Slovak, Polish, Upper Sorbian, Lower Sorbian
East Slavic: Russian, Belarusian, Ukrainian
South Slavic: Old Church Slavic, Serbian, Bulgarian, Macedonian, Slovenian
Germanic: Gothic, North Germanic, West Germanic
West Germanic: High German + English, Frisian, Dutch, Flemish, Afrikaans
High German: Old High German, German, Swiss German, Luxembourgish
English: Old English, Modern English
North Germanic: East + West Scandinavian
East Scandinavian: Danish, Swedish
West Scandinavian: Old West Norse, Icelandic, Faroese, Norwegian
Indo-Iranian: Indic + Iranian
Indic: Vedic Sanskrit, Indo-Aryan
Indo-Aryan: all the modern Indo-Aryan languages
Iranian: Avestan + the Pashto group + the Persian group + all other Iranian languages
Persian group: Persian, Tajik
Pashto group: Pashto, Waziri
Italo-Celtic: Italic + Celtic
Celtic: British Celtic, Goidelic
British Celtic: Cornish, Breton, Welsh
Goidelic: Old Irish, Modern Irish, Scots Gaelic
Italic: Latin + Romance
Romance: Sardinian + Continental Romance
Continental Romance: Eastern Romance + Italo-Western Romance
Eastern Romance: Romanian, Arumanian
Italo-Western Romance: all Romance varieties other than Sardinian and Eastern Romance
Albanian: Arvanitika, Tosk
Armenian: Classical Armenian, Modern Armenian
Modern Armenian: Adapazar, Eastern Armenian
Greek: Ancient Greek, Modern Greek

TABLE 9. Clade constraints assumed in our analyses.

and that known loanwords are unevenly distributed between ancient and nonancient languages (§2.3). We also show that overloading occurs in ancient and nonancient languages to similar degrees; this alleviates the concern that ancient language lists may be overloaded with words from too great a span of time, biasing our analyses. Finally we show that empty slots occur more often in less stable vocabulary; this is relevant to our argument that empty slots are to be avoided when assembling data sets for phylogenetic analysis (§6.1, analysis A3).

First, we measure homoplasy using the broad data set; results are shown in Table 10. In the data set, eight ancient or medieval languages are direct ancestors of thirty-nine modern languages. For each ancestral language, we find that 2–16% of the traits found in its descendants are absent in the ancestral language but present elsewhere in IE, even after excluding loans. Overall, 7% of traits in these modern languages are directly observed to be homoplastic.

Second, using the narrow data set, we divide the languages into two samples (ANCIENT languages attested over 2,000 years ago, and others) and measure the frequency of tagged loanwords in each language (Table 11, left). As expected, loanwords are tagged at a significantly lower rate in ancient languages ($p \approx 0.014$, Wilcoxon rank sum test). Borrowing rates in the later languages are consistent with those observed crosslin-

	INCLUDING LOANS			EXCLUDING LOANS		
	(a)	(b)	(c)	(d)	(e)	(f)
Ancient Greek	32	227	14.1%	24	225	10.7%
Classical Armenian	21	332	6.3%	16	323	5.0%
Latin	84	798	10.5%	64	789	8.1%
Old English	33	199	16.6%	14	174	8.0%
Old High German	46	251	18.3%	40	245	16.3%
Old Irish	34	303	11.2%	31	297	10.4%
Old West Norse	37	278	13.3%	36	277	13.0%
Vedic Sanskrit	107	1,286	8.3%	25	1,203	2.1%
Any clade	317	2,963	10.7%	200	2,873	7.0%

TABLE 10. Directly observed homoplasy in the broad data set. Shown for each ancestral language are: (a) the number of traits present in both a descendant and a nondescendant, but not in the ancestral language itself; (b) the number of traits present in any descendant; (c) their ratio; (d–f) the same after excluding recognized loanwords.

	LOANWORD RATES		AVG SLOT LOADS	
	ANCIENT	OTHER	ANCIENT	OTHER
MEAN	0.13%	2.90%	1.12	1.17
SD	0.30%	3.78%	0.10	0.12
MIN	0.00%	0.00%	1.03	1.00
MEDIAN	0.00%	1.29%	1.07	1.15
MAX	0.67%	14.10%	1.27	1.40

TABLE 11. Loanword rates (percentage of forms tagged as loans) and average slot loads (average number of forms in nonempty slots) in the narrow data set, for five ancient (before 2000 BP) and forty-seven nonancient languages. The ancient languages are Hittite, Vedic Sanskrit, Avestan, Ancient Greek, and Latin.

guistically in Bowern et al. 2011, implying that the numbers of tagged loanwords in ancient languages are indeed unrealistically small.

Third, using the narrow data set, we measure for each language the average number of forms in nonempty slots (Table 11, right). The degree of slot overloading is not significantly different in ancient and nonancient languages ($p \approx 0.42$, Wilcoxon rank sum test). This is due at least partly to the fact that the ancient-language data in IELEX originated with Ringe and Taylor’s carefully curated word lists (Ringe et al. 2002).

Finally, we use the broad data set to show that empty slots tend to occur in less stable vocabulary. For each of the 197 meaning classes, we plot the number of roots attested in it against the number of languages not attesting it (Figure 4). The former is a proxy for the lexical replacement rate in the meaning class. The correlation is significant ($p \approx 3.9 \times 10^{-6}$) and the effect size is considerable (the slope is 1.5).

5. PHYLOGENETIC METHODS.

5.1. BACKGROUND. Inferring the dates of reconstructed languages has long been associated with glottochronology, which holds that basic vocabulary replacement occurs at a universal constant rate and can be exploited to estimate the evolutionary time span between related languages (Swadesh 1952, 1955, Lees 1953). Glottochronology fell into disfavor as it became apparent that lexical replacement rates vary greatly from one language to the next, and between meaning classes. Yet the methodological crux of our work, ancestry constraints, was an element in the earliest glottochronology, which used ancestor-descendant pairs like Latin/French and Old/Modern English to measure rates of lexical replacement. Our innovation is to use ancestry constraints with a phylogenetic model; this overcomes many of the problems of traditional glottochronology. We discuss three such problems.

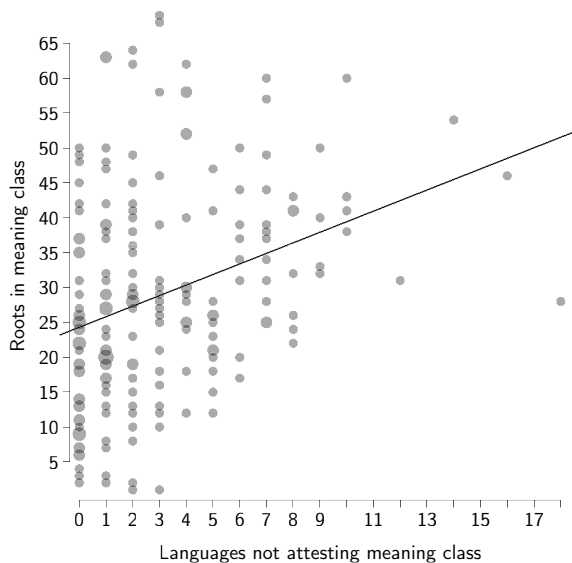


FIGURE 4. Fast-evolving meaning classes are less well attested. This is a scatterplot of all 197 meaning classes in the broad data set. Overlapping dots have been combined into larger dots.

RATE VARIATION BETWEEN LANGUAGES. The central tenet of glottochronology, that the lexical replacement rate is a universal constant, has been refuted many times. Bergsland and Vogt (1962), for example, found that Riksmål Norwegian underwent lexical replacement at roughly five times the rate of Icelandic with respect to their common ancestor, Old Norse. Blust (2000) found a similar ratio for the fastest and slowest evolving languages in a survey of 224 Austronesian languages. However, it became apparent with the availability of large word lists that such rate variation could be characterized with statistical distributions. We follow previous researchers in assuming that the rates along each branch of the IE tree obey a log-normal distribution (Bouckaert et al. 2012). This choice is driven by mathematical convenience, but it is also motivated by the fact that a log-normal distribution can be prominently peaked (implying rate homogeneity) or fairly flat (implying substantial heterogeneity) depending on its parameterization, which is not set a priori, but is inferred from the data.

MEANING-CLASS RATE VARIATION. In the ninety-four languages of our broad data set, there are over sixty roots in each of the meaning classes ‘near’, ‘hit’, ‘dirty’; and two or fewer in ‘two’, ‘three’, ‘four’. This implies a tremendous range in replacement rates across meaning classes. This has been confirmed by those working in the glottochronological tradition with Austronesian languages (Dyen et al. 1967) and those working with phylogenetic models and IE languages (Pagel et al. 2007). Both found that with lists of roughly 200 words, the lexical replacement rate was two orders of magnitude greater in the least stable meaning classes than in the most stable ones. Following a common practice in biology, we assume that rates of evolution vary from trait to trait according to a gamma distribution (Felsenstein 2004:217–20, Yang 1994). Like the log-normal distribution, the gamma distribution is mathematically convenient, and can be peaked or broad depending on its shape parameter, which can be inferred from the data.

NONINDEPENDENT DEVELOPMENT. To infer the distributions that characterize rate variation, a model should be provided with as many ancestor-descendant pairs as possible. However, such pairs tend to overlap. For example, as Swadesh (1955) noted, if both

Latin/French and Latin/Spanish were used to estimate the lexical replacement rate, the many centuries before the French-Iberian split would be counted twice. A phylogenetic model overcomes this problem by jointly inferring the tree topology, the split times, and the replacement rates along each branch. This makes it possible to aggregate information from, say, thirteen Romance languages without counting the span between Classical Latin and Proto-Romance thirteen times.

5.2. PHYLOGENETIC MODEL AND INFERENCE. Our phylogenetic analyses were carried out with a customized version of BEAST (Drummond et al. 2012), a software program for Bayesian phylogenetic analysis via Markov chain Monte Carlo (MCMC) sampling of the posterior distribution.²⁴ BEAST takes the following elements as input.

- **TRAIT DATA**, interpreted as the outcome of a diversification process. This data takes the form of a trait matrix, as described in §4.1.
- **PHYLOGENETIC MODEL**: A description of the mechanics of change whereby a single language diversifies into many. The phylogenetic model has many parts. The **TREE PRIOR** gives the probability of obtaining a tree with a specific topology and chronology, prior to seeing the data. The **TRAIT MODEL** describes how a trait evolves. The **CLOCK MODEL** describes how rates of evolution in different languages may vary.
- **CONSTRAINTS**: Specific bits of prior knowledge that shape or narrow the set of trees generated by the tree prior. **CLADE CONSTRAINTS** specify the subgroups that must be found in the inferred tree. **TIME CONSTRAINTS** specify hard or soft time intervals for splits and leaf nodes. **ANCESTRY CONSTRAINTS** force some languages to be directly ancestral to other languages.²⁵

Our analyses feature three kinds of trait models. The **RESTRICTION SITE CHARACTER (RSC)** model describes traits that can transition an arbitrary number of times between presence and absence; the frequency with which a trait is present is a global parameter (see Appendix A).²⁶ The **COVARION CHARACTER** model describes traits that can transition between being present and absent (like restriction site characters) but can also transition between being fast-evolving and slow-evolving (Appendix B). The **STOCHASTIC DOLLO CHARACTER (SDC)** model describes traits that come into existence exactly once, which suits it to traits that cannot be homoplastic (Appendix C); it is ill-suited to modeling RM traits, especially when the data set contains ancestral languages. Regardless of the trait model used, we account for the fact that a trait may exist without being observed at any of the leaves of the tree (Appendix D).

BEAST outputs a trace of first-order parameters (Tables A1–A3), which includes tree topology and chronology. The first half of the trace is discarded, and the second half is

²⁴ Our customization is viewable in a software repository (<http://github.com/whdc/ieo-beast>).

²⁵ We implemented ancestry constraints by incorporating factors into the tree prior that penalize longer lengths for the branches leading up to ancestral languages. These penalties are Gaussian functions on branch length with standard deviations of one year. Such branches do not draw a rate multiplier from the log-normal distribution for among-branch rate variation. Instead, the rate of change along these branches is set to a small value (ten transitions per million years) in order to minimize the amount by which a trait could evolve along these branches. We also added a sampling operator that attempts to move both ends of a constrained branch at once.

²⁶ This trait model is so named because it was first used to model the presence or absence, at specific sites in a genome, of DNA nucleotide sequences that are recognized by restriction enzymes (Felsenstein 1992). In the linguistic phylogenetics literature, it has also been called a two-state unequal frequencies model or a two-state continuous-time Markov chain (CTMC).

used as a POSTERIOR SAMPLE. From the trees in the sample we construct a summary tree (e.g. Figs. 1–2 above, Fig. 6 below).²⁷

5.3. EVALUATING THE STEPPE AND ANATOLIAN HYPOTHESES. For each phylogenetic analysis, we run BEAST twice: once with trait data to produce a posterior sample, and once without trait data (removing it from the XML configuration file) to produce a PRIOR SAMPLE, which is used in conjunction with the posterior sample to evaluate the steppe and Anatolian hypotheses.

To determine the degree to which the data D supports the steppe hypothesis over the Anatolian hypothesis, we calculate the following Bayes factor.

$$(6) K_{S/A} = \frac{\mathbb{P}(D \mid t_R \in \Omega_S)}{\mathbb{P}(D \mid t_R \in \Omega_A)}$$

This indicates the modeling improvement derived from constraining the root age t_R to the steppe hypothesis interval $\Omega_S = [5500, 6500]$ versus constraining it to the Anatolian hypothesis interval $\Omega_A = [8000, 9500]$. We interpret the Bayes factor conventionally: support for the steppe hypothesis is very strong if $K_{S/A} > 30$, strong if $10 < K_{S/A} < 30$, substantial if $3 < K_{S/A} < 10$, and weak to negligible if $1 < K_{S/A} < 3$ (Jeffreys 1961). In order to compute $K_{S/A}$ we apply Bayes's theorem to obtain the following.

$$(7) K_{S/A} = \frac{\Pr\{t_R \in \Omega_S \mid D\}}{\Pr\{t_R \in \Omega_S\}} \bigg/ \frac{\Pr\{t_R \in \Omega_A \mid D\}}{\Pr\{t_R \in \Omega_A\}}$$

The two numerators are estimated by noting the fraction of the posterior sample for which $t_R \in \Omega_S$ and $t_R \in \Omega_A$. The two denominators are estimated by noting the fraction of the prior sample for which $t_R \in \Omega_S$ and $t_R \in \Omega_A$.

Note that the prior sample consists of trees drawn from the tree prior, subject to the constraints. To the extent that the prior sample favors the steppe hypothesis (i.e. if the root age in this sample often falls into Ω_S), the steppe hypothesis will be penalized in the Bayes factor. This guards against concluding in favor of the steppe hypothesis on the basis of a priori assumptions about the age of the root that are implicit in the tree prior. As an additional precaution, most of our analyses use the generalized skyline coalescent tree prior (Strimmer & Pybus 2001), which gives approximately the same weight to both hypotheses.

6. EXPERIMENTS. Our phylogenetic analyses comprise four blocks:

- Block A contains our main analyses A1–3 and other analyses that give similar results despite different experimental conditions (§6.1).
- Block B contains analyses that replicate and follow up on the results of Bouckaert et al. 2012, 2013 (§6.2).
- Block C contains analyses that quantify the effects of using a single-gain trait model, of lacking ancestry constraints, and of lacking a model of among-trait rate heterogeneity (§6.3).
- Block D contains analyses that validate the time constraints that are placed on ancestral languages in our main analyses (§6.4).

²⁷ The MCMC chain length was 160 million states for analyses A1, A4–5, D1–5; 60 million states for analyses A2–3; and 20 million states for other analyses. We used a software program called TreeAnnotator (Andrew Rambaut and Alexei J. Drummond) to produce a maximum clade credibility (MCC) tree. For each tree in the posterior sample, the program assigns it a score that is the sum of the posterior probabilities of each clade in the tree. The tree with the highest score is used for the topology of the MCC tree. In plots of the MCC tree we show the median age for each node and the median branch rate multiplier for each branch, both computed from sampled trees where the leafward clade exists.

We perform a correction on the main analyses after discussing the effects of advergence on the inferred root age (§7.1).

6.1. ANALYSES A1–7: DATING PIE. This section describes our main analyses A1–3 and supporting analyses A4–7, which give similar results despite different experimental conditions. They are summarized in Figure 5. The three plots for the corrected A1–3 analyses are discussed in §7.1.

In analyses A1–3, we estimate the date of PIE by running BEAST with ancestry constraints on our narrow, medium, and broad data sets. We use the generalized skyline coalescent tree prior and the RSC trait model.²⁸ We assume log-normal distributed among-branch rate variation and gamma-distributed among-trait rate variation. We refrained from putting time constraints on unattested languages for reasons discussed in §7.2. All three data sets yield root age distributions whose median falls within the range for the steppe hypothesis, and whose 95% HPD interval does not overlap with the range for the Anatolian hypothesis. In each case the Bayes factor indicates strong or very strong support for the steppe hypothesis. The differences in the results for analyses A1–3 reflect differences in the construction of the data sets.

A1 differs from A2 in excluding modern languages that lack attested ancestors. Our intent was to minimize the effects of borrowing between, and advergent innovations within, IE clades that lack attested ancestors (British Celtic, Slavic, Albanian, Iranian, and Low German). The result is a lower inferred root age for A1 by 460 years.²⁹

A3 differs from A2 in having more languages and more meaning classes (197 rather than 143), which results in a trait matrix with many more empty slots. A language with many empty slots will seem more conservative than it actually is due to the fact that our model will underestimate the number of unique traits in the language. There are two unrelated reasons for this. First, as discussed in §4.2, empty slots tend to occur in unstable meaning classes, which contain more unique traits than stable meaning classes do. Second, as discussed in Appendix D, our model does not account for unique traits that are not directly observed. Hittite, Tocharian A, and Tocharian B are particularly sparsely attested, and are basal in the tree. Since they are treated as more conservative than they really are, the root age is driven down.³⁰

Analyses A4–7 are all based on A1, but differ in some respect. In A4, loans are excluded as described in §4.1, which increases the root age slightly. A5 operates on just the ninety-two meaning classes that Swadesh (1955) deemed most suited to glot-

²⁸ A referee asks why the RSC trait model is used in our main analyses when Bouckaert and colleagues (2013) found that the covarion trait model was measurably better. We note that the improvement with covarion was slight (a gain of 0.5% in the log marginal likelihood), that the inferred covarion parameters were not interpreted or even published, and that the covarion model in general seems to lack a cogent linguistic interpretation (§8.4). Since both RSC and covarion are essentially models of convenience, neither especially tailored to RM traits, we prefer the model that is easier to analyze. Our analysis A7 shows that using covarion does not alter our basic findings.

²⁹ Some of the difference in root ages may also arise from the fact that the trait model is somewhat ill-suited to the data, as shown in §8.1. The inferred RSC parameters (the basic rate ρ_0 and the stationary frequency of trait presence π_i ; see Appendix A) differ considerably between A1 and A2. With a better trait model we would expect these parameters to be invariant to the number of languages in the analysis.

³⁰ A referee calls our attention to a study that finds that phylogenetic methods are robust to data gaps (Wiens & Moen 2008). But that study treats only DNA characters that are never censored, whereas typically over half of the lexical traits in our analyses are unascertained. BEAST works well with either data gaps or data censoring, but not both (Appendix D). Moreover, in Wiens and Moen's simulated data, data gaps are randomly distributed without reference to other features of the data, which is not the case for lexical characters (§4.2). Finally, their study investigates the effect of data gaps on topology, not chronology.

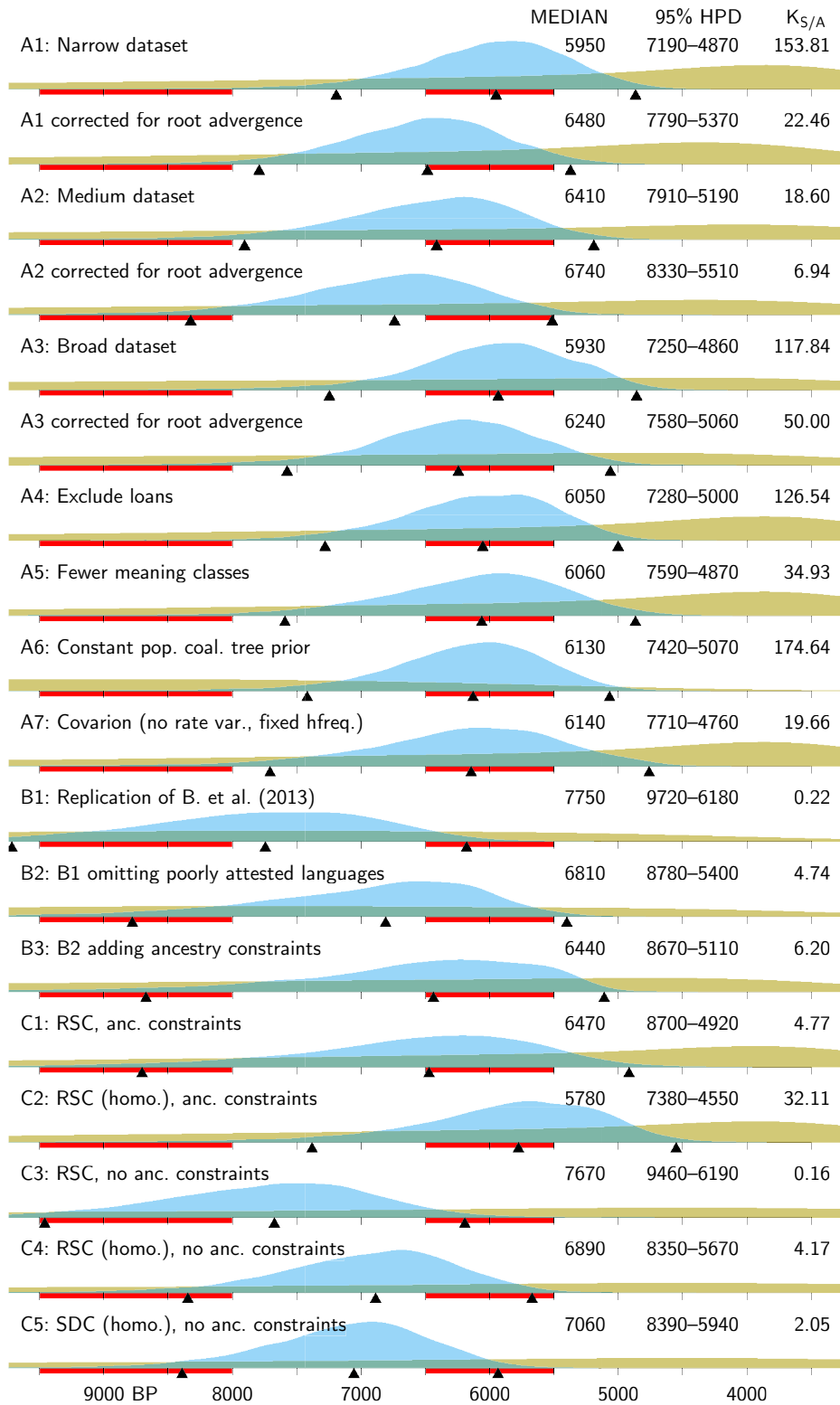


FIGURE 5. Root age prior distribution (tan) and posterior distribution (blue, rescaled by $\frac{1}{2}$) with median posterior and 95% HPD interval (arrowheads); and root age ranges for Anatolian and steppe hypotheses (red). Root age statistics and Bayes factors are printed for each analysis.

tochronological analyses. A6 has a constant population coalescent tree prior (Kingman 1982). A7 uses a covarion trait model (see Appendix B) and is otherwise identical to A1. Bouckaert and colleagues (2013) found that the covarion model was more appropriate than RSC or SDC for IELEX data. Following them, we disabled the modeling of gamma-distributed among-site rate variation and fixed the covarion parameter δ_1 to 0.5. The resulting median posterior root age is higher, and the distribution is wider.

To further validate our results, we removed the ancestry constraint for each of the eight ancestral languages in turn from analysis A1. The median root age fluctuated from 5950 to 6130 BP; the 95% HPD interval remained within the range [4850, 7480]; the Bayes factor $K_{S/A}$ fluctuated from 39 to 100.

6.2. ANALYSES B1–3: FOLLOWING UP BOUCKAERT ET AL. 2012, 2013. To ensure that our finding does not reflect differences in data sets used, we replicated work by Bouckaert and colleagues using their data set and experimental conditions. Their data set is described in §4.1. Tagged loans were excluded. Their main analysis had log-normal distributed among-branch rate variation and a covarion trait model. Among-site rate variation was not explicitly modeled, and the covarion parameter δ_1 was fixed at 0.5. Using BEAST they found a root age posterior distribution with median 7580 BP and 95% HPD interval 9350–5970. In analysis B1 we removed the phylogeographical elements from the BEAST configuration file that was published with Bouckaert et al. 2013, and replicated the inferred chronology, which supports the Anatolian hypothesis (see Fig. 5).

In analysis B2 we reran B1 after discarding the six most sparsely attested languages. The resulting root age was low enough to give significant support to the steppe hypothesis. The six languages are listed here with the number of meaning classes (out of 207) attested in each: Lycian: 34, Oscan: 52, Umbrian: 57, Old Persian: 74, Luvian: 99, Kurdish: 100. One motivation for removing these languages is that poorly attested languages bias the inferred chronology (§6.1 above). We also found that at least thirty-three of the ninety-nine Luvian words in the data set were erroneously coded for cognacy: they should have been coded as having Hittite (or in a few cases Lycian) cognates, but were coded as unique traits.³¹ The miscodings cause Luvian to seem more innovative and resulted in a higher root age in analysis B1.

In analysis B3 we reran B2 with ancestry constraints, and obtained a result that significantly favors the steppe chronology. Since the B3 trait data is the same as that of Bouckaert et al., the result cannot be attributed to IELEX contributions made by some of us. Nor, aside from the shedding of the sparsely attested languages, can it be attributed to a different choice of trait model, languages, clade constraints, or time constraints.

6.3. ANALYSES C1–5: SINGLE-GAIN TRAIT MODEL. Analyses C1–5 serve to quantify the effect of imposing the SDC trait model on the narrow data set under experimental conditions that are typical of previous work. However, a direct comparison between analysis A1 and an analysis with SDC is not possible, since (i) SDC is strictly incompatible with ancestry constraints, and (ii) as implemented in BEAST, the SDC model does not support the modeling of among-trait rate heterogeneity.

³¹ Mostly these are cases of obvious cognacy like Luvian *pāḫur* = Hittite *paḫḫur* ‘fire’. Other affected meanings are ‘I’, ‘animal’, ‘because’, ‘black’, ‘blood’, ‘bone’, ‘come’, ‘cut’, ‘die’, ‘eat’, ‘father’, ‘flow’, ‘good’, ‘head’, ‘hear’, ‘heavy’, ‘here’, ‘hit’, ‘horn’, ‘husband’, ‘lie’, ‘moon’, ‘new’, ‘seed’, ‘sit’, ‘that’, ‘this’, ‘tie’, ‘water’, ‘when’, ‘where’, and ‘year’.

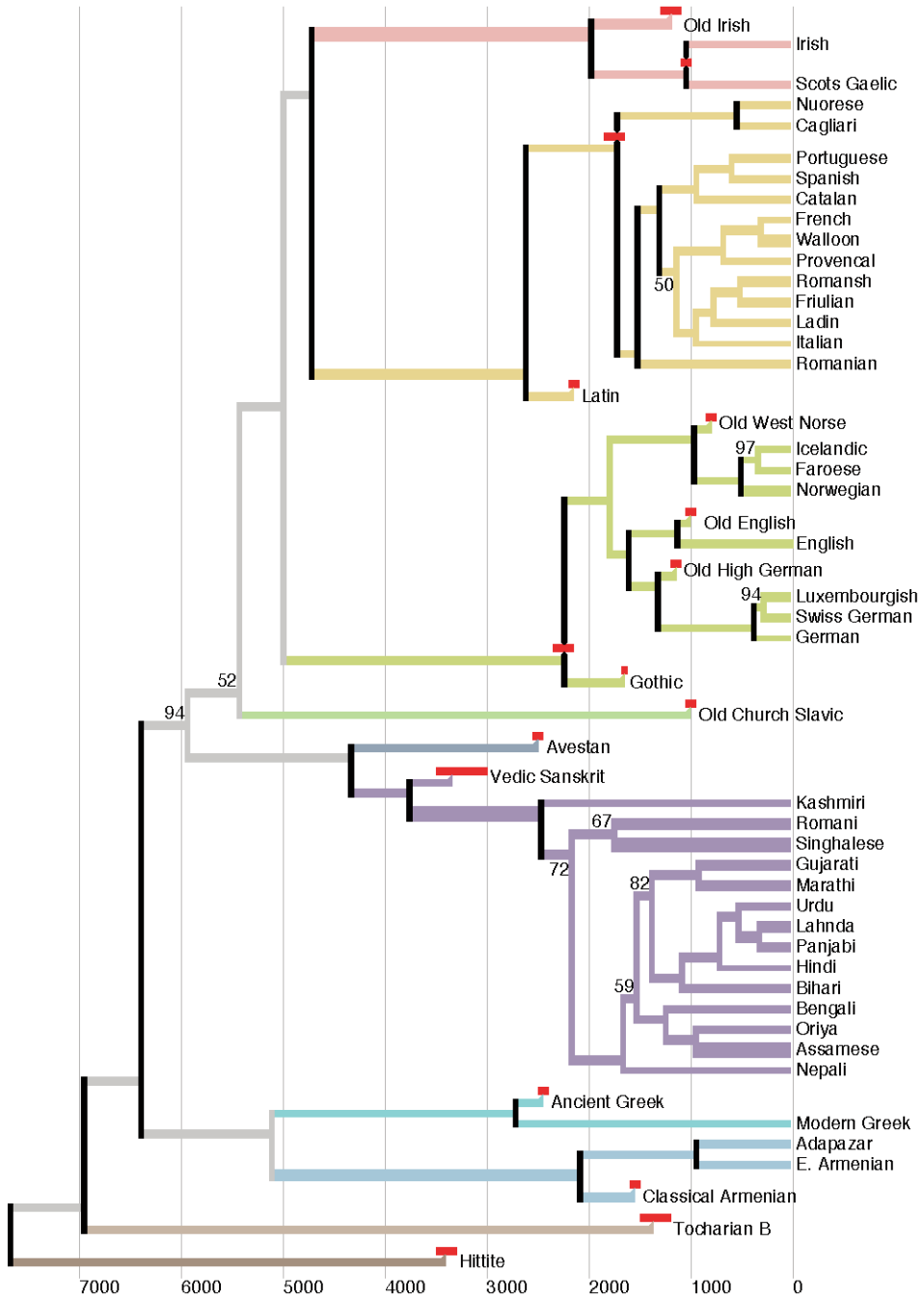
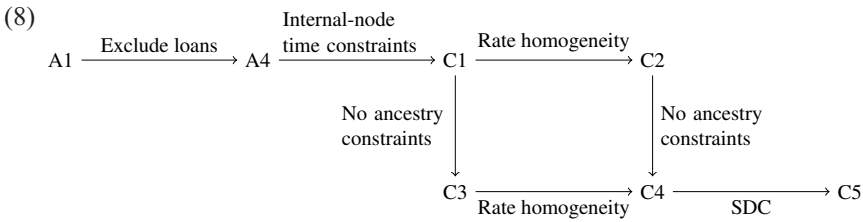


FIGURE 6. Analysis C3 summary tree. Modern languages with no ancestors in the data set are excluded, but ancestry constraints are not used. There are time constraints on splits. See Fig. 1 caption to interpret other graphical elements.

Thus we also toggle these modeling elements to control for their contributions. As in previous work, loans were excluded and some splits bear time constraints, as illustrated

in the summary tree for analysis C3 (Figure 6).³² The way in which C1–5 relate to each other, and to analysis A1 and A4, is as follows.



The results of analysis C1–5 are summarized in Fig. 5 above. C1 finds a higher root age than A4; this is an effect of time constraints on splits, as discussed in §7.2. Comparing C1 with C3 (or C2 with C4) shows that lacking ancestry constraints adds 1,200 years (or 1,110 years) to the root age. The reason for this is discussed in §8. Comparing C4 and C5 shows that using SDC adds another 170 years to the root age. Since SDC cannot posit multiple gains, it reconstructs trait gains nearer to the root than RSC would. This corresponds to a lengthening of branches that are closer to the root, with the root being placed farther in the past. Branches in leafward parts of the tree change little in length since they are subject to time constraints at leaves and splits.

Finally, we note that modeling among-site rate heterogeneity affects the root age in a systematic way. Comparing C1 with C2 (or C3 with C4) shows that modeling rate heterogeneity adds 690 years (or 780 years) to the root age. The low root age in C2 and C4 is artifactual, and C5 would find a higher root age if rate heterogeneity were modeled. We conclude this section with a mathematical description of this artifact, followed by a proof.

Under RSC, a trait that is present at time 0 will be present at time t with probability

$$(9) f_r(t) = \alpha + (1 - \alpha)e^{-rt},$$

where $r > 0$ is the rate at which the trait evolves and $\alpha \in (0, 1)$ is the stationary distribution of trait presence. (For convenience, we use a different parameterization of the model from what is found in Appendix A.) Assuming rate homogeneity, the inverse function

$$(10) f_r^{-1}(R) = -\frac{1}{r} \log \frac{R - \alpha}{1 - \alpha}$$

serves as an estimator of the evolutionary time spanned by two languages. The retention rate R is the fraction of traits of one language found in the other. Under conditions of rate heterogeneity, however, the retention rate is different from what is given in 9. With $p(r)$ as the probability distribution over rates, the retention rate between two languages is expected to be as in 11.

$$(11) E_r f_r(t) = \int_0^\infty p(r) f_r(t) dr$$

Nonetheless, an analyst may choose to ignore rate heterogeneity and apply the estimator in 10 directly to such a retention rate in order to estimate the time spanned by two languages. Using a nominal rate of evolution c , the analyst obtains 12.

$$(12) \hat{t} = f_c^{-1}(E_r f_r(t))$$

³² We used the following three date ranges based on Table 12 rather than the time constraints of Bouckaert et al. 2012, 2013: Proto-Germanic 2350–2150, Proto-Romance 1850–1650, Modern Irish/Scots Gaelic 1100–1000.

We claim that this results in progressively worse underestimates of t as t gets larger. In other words, as t grows, \hat{t}/t will fall. To see the relevance of this to dating PIE, consider that most calibration points are closer in time to the present than to PIE. Not modeling rate heterogeneity will thus shorten the rootward branches of the tree, lowering the root age.

Claim. If $\text{Var}(r) > 0$ and $t_2 > t_1$, then $\hat{t}_2/t_2 < \hat{t}_1/t_1$.

Proof. Without loss of generality, let $t_1 = 1$. Then the claim becomes

$$(13) \frac{f_c^{-1}(\mathbb{E}_r[f_r(t)])}{t} > f_c^{-1}(\mathbb{E}_r[f_r(1)]) \text{ for } t > 1.$$

Since f_c is strictly decreasing for $c > 0$, this is equivalent to the claim

$$(14) \mathbb{E}_r[f_r(t)] < f_c(t f_c^{-1}(\mathbb{E}_r[f_r(1)])) \text{ for } t > 1.$$

We define

$$(15) \phi_t(R) \equiv f_c(t f_c^{-1}(R)) = \alpha + (1 - \alpha) \left(\frac{R - \alpha}{1 - \alpha} \right)^t$$

and note that $\phi_t(R)$ does not depend on c . Thus 14 is equivalent to the claim

$$(16) \mathbb{E}_r[\phi_t(f_r(1))] < \phi_t(\mathbb{E}_r[f_r(1)]) \text{ for } t > 1.$$

Since $\phi_t(R)$ is strictly convex over the domain $R \in [\alpha, 1]$ when $t > 1$, the validity of this claim follows from Jensen’s inequality (Jensen 1906).³³

6.4. ANALYSES D1–8: ANCESTRAL LANGUAGE TIME CONSTRAINTS. Following the example of Ryder and Nicholls (2011), we perform a series of analyses to validate the time constraints on ancestral languages. Analyses D1–8 are identical to A1, except in each a different ancestral language is allowed to float: the time constraint is removed, but the ancestry constraint is left intact. The Bayes factor between \mathcal{M}_{A1} (the model in A1) and each of $\mathcal{M}_{D1}, \dots, \mathcal{M}_{D8}$ (the models in D1–8) indicates the extent to which the additional time constraint in A1 improves modeling of the data. Formally, it is a ratio of marginal likelihoods.

$$(17) K_i = \frac{\mathbb{P}(D \mid \mathcal{M}_{A1})}{\mathbb{P}(D \mid \mathcal{M}_{Di})}$$

This is equivalent to

$$(18) K_i = \frac{\text{Pr}\{t_i \in \Omega_i \mid D, \mathcal{M}_{Di}\}}{\text{Pr}\{t_i \in \Omega_i \mid \mathcal{M}_{Di}\}},$$

where t_i is the time of the floating ancestor in \mathcal{M}_{Di} , and Ω_i is the interval to which it is constrained in A1. We obtain the numerator (respectively, denominator) by noting the fraction of the posterior (respectively, prior) sample from analysis D_i for which $t_i \in \Omega_i$. The resulting Bayes factors and the prior and posterior distributions of t_i are shown in Figure 7. That they are all greater than one suggests that the time constraints (but not necessarily the ancestry constraints) are plausible in relation to the data. The median posterior root age fluctuates from 5770 BP (Latin floating) to 6120 BP (Vedic Sanskrit floating).

7. THE EFFECTS OF ADVERGENCE.

7.1. CORRECTION FOR ADVERGENCE AT THE ROOT. Readers familiar with IE languages may notice that many splits in our summary trees (Figs. 1–2) are conspicuously

³³ Jensen’s inequality is usually stated as $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$ for a random variable X and a convex function φ ; but we rely on the fact that if $\text{Var}(X) > 0$ and φ is strictly convex, then $\varphi(\mathbb{E}[X]) < \mathbb{E}[\varphi(X)]$.

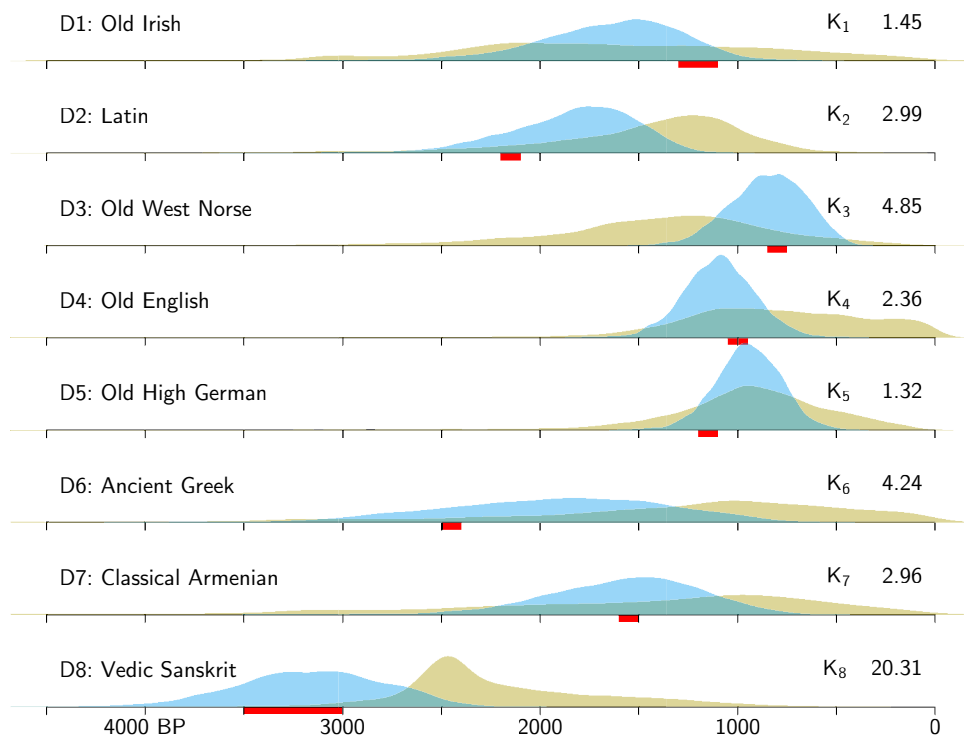


FIGURE 7. Prior (tan) and posterior (blue) distributions of the time of the floating language, and the interval it is constrained to in A1 (red). The Bayes factor K_i indicates the extent to which time-constraining the floating language in model \mathcal{M}_{D_i} improves the model with respect to the data.

younger than historical evidence allows. In Table 12 we compare eight split dates derived from historically attested events to those inferred in analyses A1–2. A2 dates are younger by 330 ± 165 years. A1 has just three of these splits since it operates on a smaller set of languages, but the split dates are on average 100 years younger still. This bias is an artifact of the interaction between advergence and the RSC trait model. Under RSC, advergent parallel gains that occur immediately after a split tend to be treated as a single gain that predates the split, which causes the ages of splits to be underestimated; this is illustrated in Figure 8. We do not know the conditions under which PIE split apart, but by extrapolating from the eight observations in Table 12, we can statistically characterize and correct for the effects of advergence associated with the first split in the tree.

Immediately after a split, advergent developments occur relatively often, because newly diverging languages share precursor traits that generate the same drift-like changes, and because they may also remain in close contact. The frequency of advergent developments tapers off as the child languages grow farther apart. To a first approximation, the node age will be underestimated by a time proportional to the number of advergent developments following the node. We assume that this perturbation of the node age has little effect on the rest of the tree since the inferred rates along adjacent branches remain accurate. (This is no longer the case if a time constraint has been placed on the split; see §7.2.) Since we observe almost no correlation between the split date and the size of the discrepancy in Table 12, we assume that advergence has tapered

SUBGROUP	DATES	NOTES AND SOURCES
Germanic	2250	By the year 100 CE, ‘the Germanic dialects had been diverging for three or four centuries’ (Jasanoff 1994); ‘a Proto-Germanic period [began] with the last centuries B.C. and [ended] in the first two centuries A.D.’ (Penzl 1988).
	A2 2060	
	A1 1960	
Romance	1750	Latin regional texts c. 250 CE (e.g. the Bu Njem ostraca) give indirect evidence that the Sardinian and non-Sardinian vowel systems had diverged structurally (J. N. Adams 2007).
	A2 1200	
	A1 1030	
Scandinavian	1500	Runic Norse ‘rather uniform across a fairly wide geographical expanse’ until this date (Fortson 2010).
	A2 1110	
Slavic	1500	‘[I]f we take into account the first ascertainable local innovations ... the linguistic evolution of Slavic after c. 500 AD signals the beginning crystallization of individual Slavic languages and language groups’ (Birnbauer 1998, similarly Timberlake 2013).
	A2 1240	
East Baltic	1300	In ‘about the seventh or eighth century A.D., ... the Letts first invaded ... the eastern part of present-day Latvia. Toponymic studies have shown that at that time there was hardly any difference between their language and that of the Lithuanians’ (Senn 1966).
	A2 970	
British Celtic	1250	‘It becomes possible to distinguish Breton and Cornish from Welsh in the course of the 8th century, when innovations arise in Welsh to differentiate it from the other two languages’ (Koch 2006, s.v. ‘Breton language’).
	A2 1152	
Modern Irish /Scots Gaelic	1050	‘[T]here is absolutely nothing to suggest that the Gaelic of Ireland, Scotland, and Man differed in any respect before the tenth century; and ... there is a body of decisive positive evidence tending to show that so far as we can tell they were identical ... [Irish and Scots] Gaelic continued to be one language ... from the tenth until the thirteenth century; but ... there are one or two significant indications, the oldest belonging to the tenth century, which point to the beginnings of the divergence between them’ (Jackson 1951; cf. Ahlqvist 1988).
	A2 490	
	A1 460	
Persian-Tajik	750	Early New Persian spoken in Central Asia (ancestral to Tajik) and Afghanistan (ancestral to Dari, i.e. Afghan Persian, not in our database) by the thirteenth century (Khanlari 1979, Windfuhr & Perry 2009).
	A2 500	

TABLE 12. Split dates for selected IE subgroups. Listed for each split is a date for the earliest evidence of linguistic differentiation, followed by dates from analyses A1–2, in years BP.

off for most of the splits in the table, and that the inferred PIE root age is underestimated by a similar amount.

To account for advergence, we augment our phylogenetic model to generate a corrected root age t_{cor} in the following way. After drawing a tree T with root age t_R from the tree prior:

- (i) Pick one of the eight (in the case of A1, three) nodes in Table 12 uniformly at random.
- (ii) Compute the discrepancy between its historically attested age and its age in T . (All relevant clades in Table 12 have been constrained to exist in T .)
- (iii) Add this discrepancy to t_R to obtain t_{cor} .

T is then used to generate the data as in any other phylogenetic model, and t_{cor} is not used for any purpose except hypothesis evaluation. Generating samples from this augmented model is simple in practice: we augment each tree in the prior and posterior samples produced by BEAST with t_{cor} according to the procedure just described. To compute $K_{S/A}$, we use the same formulas in §5.3 after substituting t_{cor} for t_R . All three corrected analyses offer at least significant support for the steppe hypothesis (see Fig. 5).

For IE, the uncorrected and corrected root age both have interpretations. The uncorrected root age corresponds to a time within the period during which pre-Proto-Anatolian and pre-PNIE were in contact, developing under the influence of advergence. The corrected root age is the analog of the historical dates in Table 12. These were de-

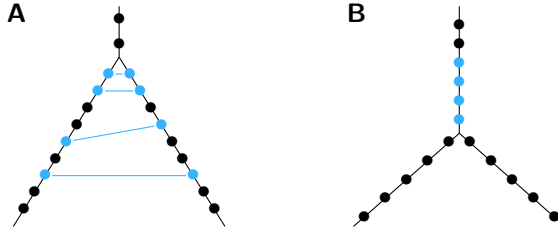


FIGURE 8. Advergence-induced inference artifact. In advergence, following a typical linguistic split (A) there are sets of homoplastic, advergent developments (linked blue dots), interspersed with nonhomoplastic developments (black dots). In the inferred tree (B), a more recent split, consistent with nonhomoplastic reconstructions of the homoplastic traits, is favored.

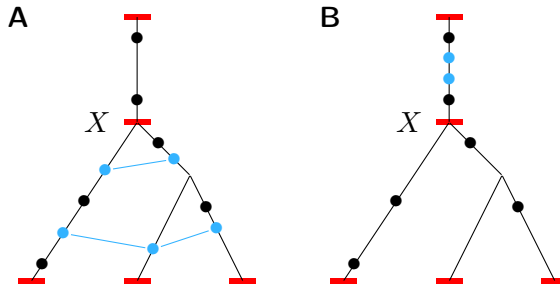


FIGURE 9. Advergence interacts with time constraints on splits. Five nodes bear time constraints (red bars); the central node *X* is an unattested language and the other four are attested languages. In the actual tree (A), there are nonhomoplastic developments (black dots) and sets of homoplastic, advergent developments (linked blue dots). How the homoplastic traits tend to be reconstructed (B) corresponds to exaggerated among-branch rate variation.

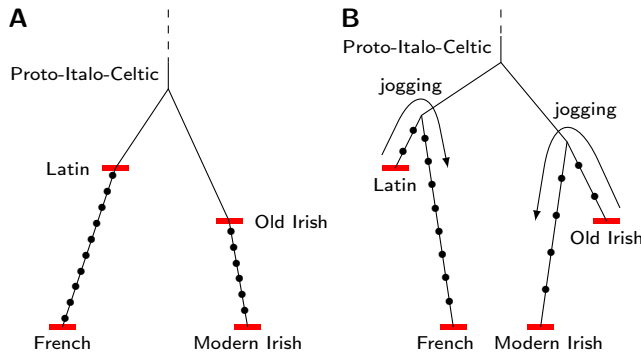


FIGURE 10. Jogging and inferred root age. In reality (A), Latin is ancestral to French and Old Irish is ancestral to Modern Irish, but in an erroneously inferred tree (B), the paths from Latin to French and from Old to Modern Irish jog rootward. An effect of jogging is that innovations between Latin and French, and between Old and Modern Irish (black dots), are reckoned as occurring over a greater evolutionary time interval; this results in a lower inferred rate of lexical change, and ultimately a higher inferred root age.

rived from either the earliest evidence of linguistic differences in the relevant diversifying populations, or the date when distinct subpopulations formed. Since both the steppe and the Anatolian hypotheses refer to the beginnings of PIE dispersal, it is better to evaluate the hypotheses using the corrected root age. The span between the corrected and uncorrected dates is analogous to the span between the first appearance of Latin re-

gional differences that were ancestral to later Romance isoglosses, around 250 CE (Table 12), and the date after which Romance varieties were not mutually comprehensible, late in the first millennium CE.³⁴

7.2. TIME CONSTRAINTS ON SPLITS. When there is advergence after a split, placing a time constraint on the split may distort the inferred rates of change on adjacent branches (Figure 9). Under RSC, parallel advergent gains resemble a single gain on the rootward side of the split, so the rates on leafward branches are underestimated, and the rate on the rootward branch is overestimated if there is an attested rootward language also constrained in time. This leads to exaggerated among-branch rate variation. If there is no attested rootward language, the overall rate of change will be underestimated, and the root will be placed farther in the past (Garrett 2006:148, n. 6). For these reasons we avoid placing time constraints on splits in most of our analyses. Analyses with predominantly modern languages (Gray & Atkinson 2003, Nicholls & Gray 2008) or predominantly ancient and medieval languages (Nicholls & Gray 2008) cannot avoid having time constraints on splits, and may thus produce inaccurate chronological results.

8. DISCUSSION AND FOLLOW-UP ANALYSES.

8.1. THE PURPOSE OF ANCESTRY CONSTRAINTS. We have shown that the ancestry relationships in Table 2 are justified by the linguistic facts (§3), and that ancestry constraints dramatically affect the root age in phylogenetic analyses (§§6.2–6.3). Now we turn to the question of why they have this effect, and the related question of why phylogenetic models typically fail to infer ancestor-descendant relationships from the data.

The answer to the first question is straightforward. Models that lack ancestry constraints will systematically analyze each ancestral language as coordinate with its descendants. Thus, for each ancestor-descendant pair, the path from one to the other jogs: it goes rootward before going leafward (Figure 10B). This jogging increases the evolutionary time separating the two languages, which leads to a lower estimate of the rate of lexical change and, ultimately, to an elongated tree. As noted in §3, the amount of jogging can be considerable. For example, in an analysis of ours that lacks ancestry constraints, jogging almost triples the evolutionary time spanned by Old and Modern Irish (Fig. 6).

This brings us to the second question: why does the model fail to infer ancestorhood, and in some cases fail so badly? At least two aspects of the phylogenetic model are to blame: the tree prior and the trait model. A realistic tree prior must assign a nonzero probability to the set of trees where one language is ancestral to another. (If an ancient language is newly discovered, it might be directly ancestral to a known language.) However, most tree priors in common use, including the generalized skyline tree prior that we and Bouckaert and colleagues (2012, 2013) use, assign an infinitesimal probability to that set of trees. Due to our Bayesian methodology, even if the maximum likelihood tree were a tree in that set, we would not find it, since the trees sampled from the posterior distribution would almost surely not be from that set.³⁵

If the tree prior explains why there is always some jogging in the inferred tree, the trait model explains why it can be so extensive. The three trait models in this work

³⁴ Clackson and Horrocks (2011:271) write that ‘speakers within the Romance area were able to communicate with each other’ throughout this period.

³⁵ There are tree priors where the marginal probability of a zero-length branch is nonzero, or, alternatively, where the sampled tree may have an internal node with one instead of two leafward branches; see Stadler 2010.

(SDC, RSC, and covarion) are poorly suited to modeling drift. As shown in §4.2, our data sets contain many STRADDLING TRAITS. These are traits that appear in the modern descendant of an attested ancestor, and also in a nondescendant of the ancestor, but not in the ancestor itself. Under the SDC trait model, a straddling trait forces the phylogenetic model to posit jogging so that the trait can be lost in the ancestral language while it is retained in the modern descendants. But even with RSC, descent relationships are hard to infer. Since precursor traits, like all RM traits, tend to be localized to one part of the tree, their corollary traits, which include straddling traits, tend to occur in closely related languages. Two closely related languages (one a descendant of an ancestral language, the other a nondescendant) will exhibit more straddling traits than RSC can account for, unless jogging is introduced into the tree. This account of jogging is supported by two experiments: §8.2 shows that the number of straddling traits in closely related languages is higher than predicted when ancestry constraints are in place, and §8.3 shows that an elevated number of such straddling traits produces jogging when ancestral constraints are removed. We have not probed the behavior of the covarion trait model as thoroughly, but there are theoretical reasons to believe that it is also poorly suited to modeling drift, as discussed in §8.4.

The foregoing implies that ancestry might be inferable with an appropriate tree prior and trait model, though drift might still produce data with insufficient statistical information for inferring descent relationships. If descent relationships are known from sources outside the data, it is best to use ancestry constraints.

8.2. STRADDLING TRAITS ANALYSIS. In this analysis and those of §8.3, we construct triplets of languages and analyze each triplet in isolation. Each row in Table 13 corresponds to a triplet, which consists of an ancestral language A , a descendant B , and a close nondescendant C .

A	B	C	N	n	\tilde{n}	t_A	$t_J^{(1)} - t_A$	$t_J^{(2)} - t_J^{(1)}$
OIr	Scots Gaelic	L	334	1	0.410	1200	+432	+20
OIr	Irish	L	338	3	0.415	1200	+468	+18
L	Catalan	OIr	377	1	0.620	2150	+77	+52
L	Spanish	OIr	339	2	0.557	2150	+0	+0
OWN	Icelandic	OHG	234	2	0.150	800	+32	+15
OWN	Norwegian	OHG	241	6	0.154	800	+203	+14
OE	English	OHG	213	3	0.117	1000	+93	+20
OHG	Swiss German	OE	216	1	0.128	1150	+53	+28
OHG	Luxembourgish	OE	217	3	0.129	1150	+116	+20
Grk	Modern Greek	Arm	342	2	0.584	2450	+0	+15
Arm	E. Armenian	Grk	348	0	0.486	1550	+154	+36
Arm	Adapazar	Grk	312	0	0.436	1550	+108	+46
Skt	Singhalese	Av	332	1	0.477	3250	+546	+32
Skt	Kashmiri	Av	324	4	0.466	3250	+485	+50

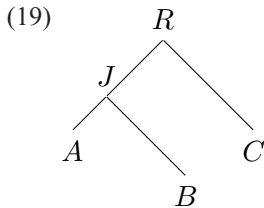
TABLE 13. Analysis of drift and jogging on triplets of languages. For each ancestral language A , B is a descendant and C is a close nondescendant. There are N traits attested in one or more of $A-C$ (loanwords excluded); of these, n is the number of traits straddling A ; \tilde{n} is the expected number of such traits based on parameters inferred in analysis A4; t_A is the center of the time interval to which A is constrained; $t_J^{(1)}$, $t_J^{(2)}$ are values for the TMRCA of A and B under scenarios described in the text. Abbreviations: Arm: Classical Armenian, Av: Avestan, Grk: Ancient Greek, L: Latin, OE: Old English, OHG: Old High German, OIr: Old Irish, OWN: Old West Norse, Skt: Vedic Sanskrit.

In this analysis we investigate the discrepancy between the actual number of straddling traits and the expected number of traits under the phylogenetic model of analysis A4, in which ancestry constraints are used. If, as we theorize, precursor traits condition the gain of corollary traits, then parallel gains would tend to occur in closely related languages. We would expect the number of straddling traits in *B* and *C* to be higher than predicted by A4. In this analysis, as in A4, tagged loanwords are excluded so as to rule out borrowing as a confounding factor.

For each triplet *A–C*, we consider the *N* traits in the data set of A4 that appear in *A*, *B*, or *C*, after excluding meaning classes in which *A*, *B*, or *C* has an empty slot. Of these traits, we count the number of straddling traits: traits present in *B* and *C*, but not in *A*. In most cases this number *n* is greater than the expected number \tilde{n} . The latter is computed by constructing a rooted tree of the three languages, with the topology and chronology as they are in the summary tree of analysis A4. First-order parameters are set to their median posterior values: $\pi_1 = 0.0103$, $\rho_0 = 7.66 \times 10^{-6}$, $\sigma = 0.351$, $\alpha = 1.004$. With branch and trait rate multipliers integrated out, we compute the fraction of observed traits that are expected to be straddling, which is scaled by *N* to obtain \tilde{n} . For ancestral languages with one or two descendants, Table 13 lists all triplets; otherwise it lists the two triplets with the lowest and highest *n/N* ratio.

Except when the ancestral language is Classical Armenian, we find that the number of straddling traits is greater than expected, which indicates an elevated number of parallel gains between closely related languages. This discrepancy is greatest of all for Germanic triplets, where *A* and *C* are especially close, and advergence between the descendants of *A* and *C* boosts the number of straddling traits.

8.3. MAXIMUM LIKELIHOOD ANALYSES. In these analyses we continue to analyze the triplets of languages in Table 13 as we investigate the impact of straddling traits on jogging. We expect that jogging will increase with the number of straddling traits. We begin by constructing a rooted tree with the following topology.



We set the times of the leaves to the center of the time interval to which they are constrained in analysis A4, and infer the split times via maximum likelihood. The data consists of seven counts, one for each pattern of attestation (three counts for traits attested exclusively in one language; three counts for traits attested in two languages; a count for traits attested in all languages), summing to *N*. Note that one such count denotes the number of straddling traits. Branch and trait rate multipliers are integrated out. This analysis is done twice.

- (i) First-order parameters are set to their median posterior values in analysis A4, as described in §8.2.
- (ii) The same analysis is repeated with the count for straddling traits increased by one.

In Table 13, $t_J^{(i)}$ refers to the time of node *J* in scenario *i*. The amount of jogging in scenario *i*, or the length by which the path between *A* and *B* increases, is thus $2(t_J^{(i)} - t_A)$. Under the first scenario, we find that the number of traits straddling an ancestor corre-

lates poorly with the amount of jogging, which depends also on the other pattern counts. However, we find that adding a straddling trait, as in the second scenario, causes the amount of jogging to increase substantially unless the pattern counts are such that the amount of jogging is pinned at zero, as is the case for the Latin/Spanish/Old Irish triplet.

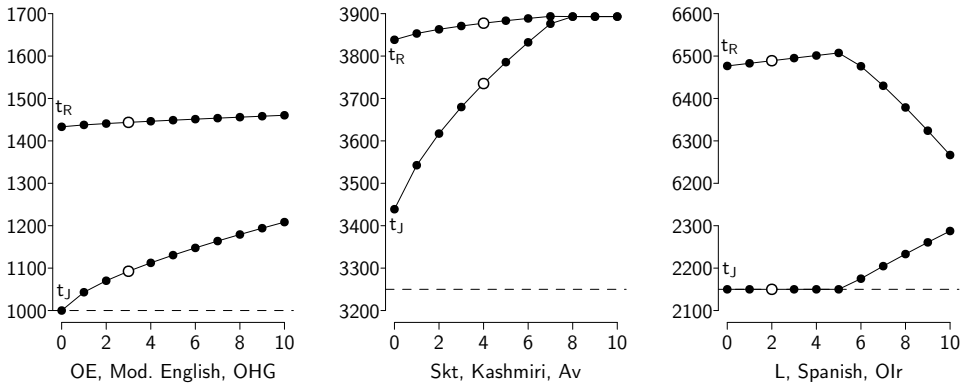
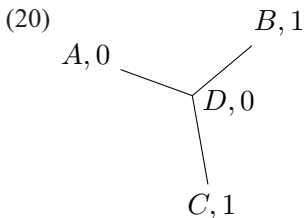


FIGURE 11. Jogging vs. number of straddling traits. For three selected triplets, we plot t_R and t_J as the number of straddling traits is varied between zero and ten. Clear dots indicate the actual number of straddling traits in our data set. The dashed line indicates t_A .

To enlarge on the relationship between jogging and the number of straddling traits, we vary the number of straddling traits and plot the inferred t_J and t_R for three selected triplets (Figure 11). We observe that t_J rises with the number of straddling traits, unless it is pinned at zero, as is the case for the first many points in the third plot. The inferred t_R is realistic in the first two plots, but not in the third. In all cases, however, the effect of straddling traits on t_R is relatively small in the vicinity of the actual number of straddling traits observed.

8.4. PROPERTIES OF COVARION TRAITS. When we compare the mathematical properties of covarion traits to the hypothesized behavior of RM traits, two incongruities stand out. First, the covarion model describes a time-reversible trait, whereas in the general case, a precursor trait can give rise to a corollary trait, but not vice versa. (Most types of semantic change are unidirectional.) For a further illustration of this point, consider the following unrooted tree with equal-length branches, labeled with observed values of an RM trait at languages A – D .



When a time-reversible trait is made to evolve along a tree with a strict clock, the root can be placed anywhere on the tree without altering the tree’s likelihood (Felsenstein 2004:204–5). But our account of drift (and basic linguistic intuition) suggests that it is more likely for the tree to be rooted at A than at B . If rooted at A , the parallel gain in B and C can be explained as drift. If rooted at B , then its loss in D and reappearance in C require two separate explanations.

A second mismatch between the covarion model and RM traits concerns the fact that the latter tend to be localized: generally a trait can be gained only where its precursor trait is present. For instance, consider traits that are absent in OE and present in Modern English. We would expect to see more such traits in Old High German than in Hittite. This is borne out in IELEX data: of the thirty-five traits that are absent in OE but present in Modern English as nonloanwords, three are also in Old High German ('man', 'blow', 'fat') and none are in Hittite. The covarion trait model, however, would predict the opposite: unless a language descends from OE, the farther it is from OE, the more such traits it is expected to have. We conclude this section by sketching a mathematical proof for a more general form of this claim.

Claim. Let $(X_t)_{t \geq 0}$ be a continuous-time Markov chain (CTMC) that represents a covarion trait, with states C0, C1, H0, H1 as defined in Appendix B and a transition rate matrix

$$(21) \quad A = \begin{matrix} & \begin{matrix} \text{C0} & \text{C1} & \text{H0} & \text{H1} \end{matrix} \\ \begin{matrix} \text{C0} \\ \text{C1} \\ \text{H0} \\ \text{H1} \end{matrix} & \begin{pmatrix} -a\pi_1 - s\delta_1 & a\pi_1 & s\delta_1 & 0 \\ a\pi_0 & -a\pi_0 - s\delta_1 & 0 & s\delta_1 \\ s\delta_0 & 0 & -\pi_1 - s\delta_0 & \pi_1 \\ 0 & s\delta_0 & \pi_0 & -\pi_0 - s\delta_0 \end{pmatrix} \end{matrix}$$

where $\pi_0, \pi_1, \delta_0, \delta_1, a, s > 0$ and $\pi_0 + \pi_1 = 1$ and $\delta_0 + \delta_1 = 1$. If the trait is absent at time zero, the probability that the trait is present at time t grows monotonically with t . In other words, $\Pr\{X_t \in \{C1, H1\} \mid X_0 = C0\}$ and $\Pr\{X_t \in \{C1, H1\} \mid X_0 = H0\}$ are both nondecreasing with respect to t .

Sketch of proof. For convenience we define a second CTMC $(Y_t)_{t \geq 0}$ that indicates whether the trait is hot or cold: $Y_t = 1$ if $X_t \in \{H0, H1\}$, or else $Y_t = 0$. Note that Y is time-homogeneous with the transition rate matrix in 22.

$$(22) \quad B = s \begin{pmatrix} -\delta_1 & \delta_1 \\ \delta_0 & -\delta_0 \end{pmatrix}$$

We also define a family of random variables $(Z_t)_{t \geq 0}$ that indicates whether the trait is present: $Z_t = 1$ if $X_t \in \{C1, H1\}$, or else $Z_t = 0$. Conditioned on Y , Z is a nonhomogeneous CTMC whose transition matrix is given in 23.

$$(23) \quad C_t = [Y_t + a(1 - Y_t)] \begin{pmatrix} -\pi_1 & \pi_1 \\ \pi_0 & -\pi_0 \end{pmatrix}$$

Thus, conditioned on Y and on the trait being absent at time 0, the trait is present at time t with probability

$$(24) \quad \Pr\{Z_t = 1 \mid Z_0 = 0, Y\} = \pi_1 - \pi_1 \exp\left(-\int_0^t Y_s + a(1 - Y_s) ds\right),$$

which is a nondecreasing function of t . Removing the conditioning on Y involves summing over the possible paths of Y in the time interval $[0, t]$. Regardless of how this is done, the summands are all nondecreasing functions of t , and thus the sum must also be a nondecreasing function of t .

Remark. We can apply this to the example of a trait that is missing in OE and present in Modern English by saying that X_0 is the state of the trait in OE, that X_t is the state of the trait in some nondescendant of OE, and that t is the evolutionary time spanned by these two languages. Since this trait is present in Modern English, we know the distribution of X_0 . The probability that the trait is present in a nondescendant $\Pr\{Z_t = 1 \mid Z_0 = 0\}$ grows monotonically with t .

Remark. An RSC trait is equivalent to a covarion trait with $a = 1$. Thus the foregoing holds for RSC traits as well.

8.5. IMPLICATIONS OUTSIDE OF INDO-EUROPEAN. For IE, with a long and rich linguistic history, we have shown that it is desirable to model known relationships between ancestor and descendant languages. This raises obvious questions for language families with few or no documented ancestor languages. Such families are the norm, and from the earliest days of lexicostatistics linguists have hoped to understand their prehistory through statistical analysis of linguistic data. We share the view that methods like those we and others use can answer questions about chronology. Such studies have been done elsewhere, in fact earlier for Austronesian (Gray & Jordan 2000) than for IE (Gray & Atkinson 2003). Using methods similar to those we criticize for IE, this Austronesian work (Greenhill & Gray 2005, 2009, Gray et al. 2009, Greenhill et al. 2010) has reached results that we understand to support the consensus of archaeologists and historical linguists. It is natural to wonder about the difference.

A variety of factors could be responsible for differences between Austronesian phylogenetic analyses (where we have no reason to doubt the inferred chronology) and previous IE analyses. First, differences both in the nature of Austronesian and IE documentation and in the history of scholarship over the last two centuries make it far less likely that related words in any two IE languages have escaped notice. For example, the PIE root **h₂enh₁-* ‘breathe’ has derivatives in words like Persian *jānvār* and Scots Gaelic *ainmhidh*, both meaning ‘animal’; in each word the only trace of that root is in the first vowel and *n*. Without information about early Indo-Iranian and Celtic, an analyst might not recognize this shared RM trait. It seems possible that distant homoplasy obscured by phonological and morphological change is less often recorded in the Austronesian data set.

Second, Austronesian and IE ‘roots’ are not the same linguistically, and may generate different patterns of derivational or semantic drift. We have tried to understand the evolutionary properties of RM traits in IE; we do not know whether Austronesian traits have a significantly lower incidence of drift-induced homoplasy. Third, the time constraints on early Austronesian splits are tighter than on early IE splits; this may have constrained chronological inference in a beneficial way.³⁶ Fourth, the Austronesian and IE data sets are different; even a small number of errors can make a significant difference. And finally, we would not exclude the possibility of fortuitous interactions between data gaps, trait models, and other components of phylogenetic analysis. We do not know enough to say which of these (or other) factors play a role. We simply note that the principles behind a crosslinguistically reliable toolkit for inferring relationships and chronology are underdeveloped, and we hope that future work can reconcile the differences noted here.

9. CONCLUSIONS. Our most important conclusion is that statistical phylogenetic analysis strongly supports the steppe hypothesis of IE origins, contrary to the claims of previous research. This in turn contributes to the study of Eurasian linguistic prehistory, indicating that IE language dispersal was not driven by the spread of agriculture.

Our work made crucial use of documented ancestor-descendant relationships to estimate the rate of lexical trait evolution. Descent relationships can be obscured by drift in

³⁶ Gray and colleagues (2009) constrained Proto-Oceanic to 3600–3200 BP, whereas Bouckaert and colleagues’ (2012, 2013) earliest constraint with a tight upper bound is on Proto-Balto-Slavic, which has a much more relaxed constraint of 3400–2000 BP. The earliest IE split with comparably tight bounds is Proto-Northwest-Germanic, which has bounds of roughly 2000–1750 BP. Note also that Proto-Oceanic is ancestral to roughly half of the languages in the Austronesian analyses. For comparison, Proto-Balto-Slavic and Proto-Northwest-Germanic are each ancestral to a sixth of the languages in the IE analyses.

lexical traits, and are hard to infer with the trait models and tree priors commonly used in Bayesian analysis. We thus constrained the IE tree to reflect uncontroversial ancestry relationships. Closely related languages that remain in contact can evolve in parallel, biasing chronological inference, but even when we corrected for this effect, we found robust support for the steppe hypothesis.

To defend the Anatolian hypothesis in light of these results, it would be necessary to challenge the reality that underlies ancestry constraints. For example, given a homoplastic trait like [**pod-*, ‘leg’] that is present in Modern Greek and modern IA languages, but not their ancient ancestors, one could claim that the PIE word for ‘foot’ also meant ‘leg’ in a common ancestor of Greek and IA, but that this meaning is undocumented in the copious textual record of Ancient Greek and Sanskrit. Such a claim is unsupported; it is more sensible to assume that a typologically common shift happened independently in languages with the ‘leg’ meaning. Our approach, which is based on the findings of diachronic semantic typology and IE lexicography, does not require positing unobserved dialects and word meanings in otherwise well-documented languages.

We have also made methodological contributions to the field of linguistic phylogenetics. We introduced ancestry constraints, which may be useful in analyzing other language families with a long written history (e.g. Afro-Asiatic and Sino-Tibetan). For analyzing lexical data in any family, we highlighted the distinction between cognate and root-meaning traits. The former correspond to traditional units of analysis in historical linguistics, while the latter (derived from basic vocabulary lists) are more often analyzed in statistical phylogenetics and have different evolutionary properties. We hope that many kinds of linguistic traits will be studied in future phylogenetic research, but it is important to understand the patterns of change that are characteristic of each of them.

Because previous statistical phylogenetic research supported the Anatolian hypothesis, linguists who find that hypothesis implausible for other reasons may dismiss statistical analyses that purport to determine ancestral chronology. For example, one recent IE textbook writes that ‘the jury is still out on whether phylogenetic dating can help solve the problem of how old the IE language family is’ (Clackson 2007:19); another does not mention phylogenetic arguments at all in discussing the Anatolian hypothesis (Fortson 2010). Both are excellent books whose aporia, we speculate, reflects uncertainty about a new method whose results may have seemed implausible, perhaps combined with the lingering bad odor of glottochronology. A final conclusion of our work is therefore that statistical phylogenetic analysis can yield reliable information about prehistoric chronology, at least where all of the available data is taken into consideration. Desiderata for future work include a stochastic model of lexical trait evolution that accounts for the effect of precursor traits and for advergence in closely related languages, to be evaluated against existing models. Meanwhile, we hope our results contribute to a synthesis of methods from linguistics and other disciplines seeking to understand the prehistoric spread of languages and their speakers.

APPENDICES: PHYLOGENETIC MODEL

APPENDIX A: FULL MODEL

Our phylogenetic model is a patchwork of many elements. To save the reader the trouble of tracking down their mathematical descriptions, or of reading BEAST source code, we give an overview of the model behind analyses A1–3. Table A1 lists the first-order parameters of the model. A priori, these parameters are independent of one another.

Parameter T is a binary tree with L leaves, where L is the number of languages, both ancestral and nonancestral. Each node has a time value that is not greater (not older) than its parent’s time, unless the node is an ancestral language, in which case it has the same time value as its parent. T is drawn from a distribution with the unnormalized density in A1.

PARAMETER	BEAST XML	SYMBOL	UNIT	PRIOR
Tree topology & chronology		T		See text
Total number of traits		S		$p(S) \propto S^{-1}$
Base rate	uclid.mean	ρ_0	yr^{-1}	$p(\rho_0) \propto \rho_0^{-1}$
Shape for among-branch var.	uclid.stdev	σ		$\sigma \sim \text{Exp}(2.5)$
Shape for among-site rate var.	alpha	α		$\alpha \sim \text{Exp}(2.5)$
Stationary frequency at 1	frequencies2	π_1		Flat on (0, 1)

TABLE A1. First-order parameters in a phylogenetic model (RSC trait model).

$$(A1) \quad p(T) \propto f_0(T) \cdot f_1(T) \cdot f_2(T) \cdots f_C(T)$$

The factor $f_0(T)$ refers to a marginal distribution of an ordinary tree prior for L languages. In this marginal distribution, the time of each ancestral language is equal to the time of its parent (see n. 25 for a description of the implementation). The factors $f_1(T), f_2(T), \dots, f_C(T)$ represent constraints on the tree. Each constraint $f_c(T)$ involves two elements: a set of languages \mathcal{A}_c and a function $g_c: (-\infty, \infty) \rightarrow [0, \infty)$, so that:

$$(A2) \quad f_c(T) = \begin{cases} g_c(\text{TMRCa of } \mathcal{A}_c \text{ in } T) & \text{if } \mathcal{A}_c \text{ is monophyletic in } T, \\ 0 & \text{otherwise.} \end{cases}$$

When g_c is a positive constant, the constraint merely ensures that a clade for \mathcal{A}_c is present in T . When g_c is a nonconstant function, it can be used to ensure that the TMRCa of \mathcal{A}_c falls in or near a specific time interval. When \mathcal{A}_c contains just one language, the constraint acts to bound the time of a leaf node. Constraints are also used to ensure that ancestral languages are positioned correctly in the tree. Let A be an ancestral language and let \mathcal{D}_A be the set of A 's descendants. We use one constraint to ensure that \mathcal{D}_A is a clade, and a second constraint to ensure that $\{A\} \cup \mathcal{D}_A$ is a clade. The latter constraint can also be used to bound A in time.

In the universe of the model there are S traits, not all of which are observed. Each trait is present at the root with probability π_1 . As trait i goes along branch j , it evolves according to the instantaneous rate matrix

$$(A3) \quad \frac{\rho_{ij}}{2\pi_0\pi_1} \cdot \begin{pmatrix} -\pi_1 & \pi_1 \\ \pi_0 & -\pi_0 \end{pmatrix},$$

where $\pi_0 = 1 - \pi_1$ and ρ_{ij} is the expected number of transitions per unit time for trait i on branch j . In a small amount of time Δt , the probability of transitioning is $\rho_{ij}\Delta t/(2\pi_0)$ if the trait is absent (0) and $\rho_{ij}\Delta t/(2\pi_1)$ if the trait is present (1). The rate ρ_{ij} is the product of three model parameters.

$$(A4) \quad \rho_{ij} = \rho_0 \cdot s_i \cdot b_j$$

The base rate ρ_0 is the mean rate at which traits evolve; it is a first-order parameter. Each trait multiplier s_i is drawn from a gamma distribution with a mean of one.

$$(A5) \quad s_i \sim \Gamma(\alpha, 1/\alpha)$$

Each branch multiplier b_j is drawn from a log-normal distribution with a mean of one.

$$(A6) \quad b_j \sim \log \mathcal{N}(-\sigma^2/2, \sigma^2)$$

The trait multiplier s_i is different for each trait but the same for all branches, while the branch multiplier b_j is different for each branch but the same for all traits.³⁷

After all S traits have been evolved to the leaves, those that are 0 at all leaves are discarded. N traits remain. These are formed into an $L \times N$ binary matrix, y . The cell in the l th row and n th column (y_{ln}) is 1 if language l has trait n , or else 0. Hidden cells are generated by positing another $L \times N$ binary matrix q independently of the rest of the model. The prior distribution of q is unknown. The data x , which is an $L \times N$ matrix with three possible values in each cell, is generated deterministically from y and q : a cell x_{ln} is marked as hidden ('?') if $q_{ln} = 1$, or else $x_{ln} = y_{ln}$.

At this point, a careful reader will notice that this generative process could result in a column without ones in x , which implies that a trait can enter the data without being observed in any language. This issue is taken up in Appendix D.

APPENDIX B: COVARION TRAITS

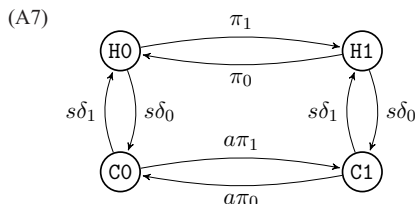
Some of our analyses use BEAST's binary covarion trait model instead of restriction site characters. This adds three first-order parameters (Table A2). We write a for the parameter bcov.alpha, to avoid confusing it with the shape parameter for gamma-distributed rate variation. For convenience we define $\delta_0 = 1 - \delta_1$ and, as before, $\pi_0 = 1 - \pi_1$.

³⁷ During inference, BEAST approximates the gamma distribution for among-trait variation using a discrete distribution with a small number of point masses (in our case, four). This is a common practice (Yang 1994). The log-normal distribution for among-branch variation is called the UCLD (uncorrelated log-normal distributed) relaxed clock model (Drummond et al. 2006).

PARAMETER	BEAST XML	SYMBOL	PRIOR
Stationary frequency of hot states	hfrequencies2	δ_1	Flat on (0, 1)
Rate reduction for cold states	bcov.alpha	a	Flat on (0, 1)
Transition rate between hot and cold	bcov.s	s	$s \sim \Gamma(0.05, 10)$

TABLE A2. Additional first-order parameters for a covarion trait model. See also Table A1.

In the binary covarion model, there are two states for trait presence, a ‘hot’ state (H1) and a ‘cold’ state (C1). Similarly, there are two states for trait absence, H0 and C0. The transition rates between cold states are a fraction a of the transition rates between hot states, as diagrammed in A7.



The transition rates are set so that, a priori, the expected fraction of traits that are hot at any point in the tree is δ_1 , and the expected fraction of traits that are present is π_1 . In BEAST, the rates are normalized so that the expected number of transitions between presence (of either kind) and absence (of either kind) is ρ_{ij} , just as with restriction site characters. This results in the instantaneous rate matrix in A8.

(A8)

$$\frac{\rho_{ij}}{2(a\delta_0 + \delta_1)\pi_0\pi_1} \cdot \begin{pmatrix} & \text{C0} & \text{C1} & \text{H0} & \text{H1} \\ \text{C0} & -a\pi_1 - s\delta_1 & a\pi_1 & s\delta_1 & 0 \\ \text{C1} & a\pi_0 & -a\pi_0 - s\delta_1 & 0 & s\delta_1 \\ \text{H0} & s\delta_0 & 0 & -\pi_1 - s\delta_0 & \pi_1 \\ \text{H1} & 0 & s\delta_0 & \pi_0 & -\pi_0 - s\delta_0 \end{pmatrix}$$

APPENDIX C: STOCHASTIC DOLLO CHARACTERS

Some of our analyses use the SDC trait model instead of RSC. These analyses are parameterized differently, with first-order parameters as listed in Table A3.

PARAMETER	BEAST XML	SYMBOL	UNIT	PRIOR
Tree topology & chronology		T		See Appendix A
Global birth rate		λ	yr ⁻¹	$p(\lambda) \propto \lambda^{-1}$
Per-capita death rate	sdc.death.rate	μ	yr ⁻¹	$p(\mu) \propto \mu^{-1}$
Shape for among-branch rate var.	uclid.stdev	σ		$\sigma \sim \text{Exp}(2.5)$

TABLE A3. First-order parameters for a phylogenetic model (SDC trait model).

The base rate ρ_0 and the stationary frequency of trait presence π_1 are replaced by a global birth rate λ and a per-capita death rate μ . Since SDC posits an infinite number of traits (only a finite subset of which are present at any of the leaves), the parameter S is no longer necessary. The shape parameter for the gamma distribution α is also gone because, in BEAST, among-site rate variation is not supported in the SDC trait model.

With SDC, the tree obtained from the tree prior is augmented with an extra branch that extends from the root to a node positioned at infinity (eternity past). The branch multiplier for this branch is one. For other branches it is drawn from a log-normal distribution, as described in Appendix A.

The global birth rate is the rate at which traits are born on the augmented tree. Over a duration Δt on branch j , the mean number of traits born in that stretch of the tree is $\Delta t b_j \lambda$. After a trait is born, it dies at the instantaneous rate of $b_j \mu$, where b_j is the branch multiplier of whatever branch the trait is on. Over a short duration Δt , the probability of the trait dying on branch j is $\Delta t b_j \mu$. Though an infinite number of traits are born in the augmented tree, only a finite number N survive to any of the leaves. These N traits are formed into the $N \times L$ matrix y . The data x is generated from y via the process described in Appendix A.

The SDC model as implemented in BEAST hews closely to the original description of SDC by Nicholls and Gray (2008), except that λ is integrated out in order to speed up inference (Alekseyenko et al. 2008). It lacks the model of lexicographic coverage proposed by Ryder and Nicholls (2011).

APPENDIX D: ASCERTAINMENT BIAS CORRECTION

The following discussion applies to the RSC or covarion trait model. For an account of the ascertainment model used with SDC, see Nicholls & Gray 2008. Ascertainment bias correction is how BEAST accounts for the possibility that some traits are unobserved at all leaves in the tree, and thus unascertained, that is, censored in the data.

A Bayesian account of BEAST's ascertainment model requires us to posit S traits, of which N are observed. We write q_n for the probability that a trait will have the outcome of the n th observed trait, conditioned on the first-order parameters given in Table A1 or A2, and on the branch rate multipliers b . Let q_0 be the probability that a trait will not be observed at the leaves of the tree, conditioned on the same things. Then, the probability of the data D is as in A9.

$$(A9) \quad \mathbb{P}(D | S, \dots) = \binom{S}{N} q_0^{S-N} \prod_{n=1}^N q_n \text{ for } S \geq N$$

S is a nuisance parameter and can be summed out. Using an improper prior of $p(S) = 1/S$, one obtains an unnormalized likelihood function as in A10.

$$(A10) \quad \sum_{S=N}^{\infty} \frac{\mathbb{P}(D | S, \dots)}{S} = \frac{1}{N} \prod_{n=1}^N \frac{q_n}{1 - q_0}$$

Conditioning on N then causes the factor of $1/N$ to fall away, giving a normalized likelihood function.

$$(A11) \quad \mathbb{P}(D | N, \dots) = \prod_{n=1}^N \frac{q_n}{1 - q_0}$$

This was what was proposed by Felsenstein (1992) as a method to account for unascertained traits; it is also what is implemented in BEAST.

In BEAST, however, this method has not been extended to work with hidden cells. When a language is sparsely attested, the probability of an unascertained trait (q_0 above) ought to grow, since there are now two reasons for a trait to be unascertained: it could be truly absent from all languages; or it could be unattested due to poor lexicographic coverage. Not accounting for the latter will result in sparsely attested languages being treated as more conservative than they are. As an illustration of this effect, suppose that there is a language A that is unattested in 100 out of a total of 200 meaning classes, and that A attests ten traits that no other language has. In reality, A must have roughly twenty unique traits, half unattested, but since this possibility is ignored, A is treated as having only ten unique traits, and thus as more conservative than it is.

REFERENCES

- ADAMS, DOUGLAS Q. 2013. *A dictionary of Tocharian B, revised and greatly enlarged*. 2nd edn. Amsterdam: Rodopi.
- ADAMS, J. N. 2007. *The regional diversification of Latin 200 BC–AD 600*. Cambridge: Cambridge University Press.
- ADAMS, J. N. 2013. *Social variation and the Latin language*. Cambridge: Cambridge University Press.
- ADJARIAN, H. 1909. *Classification des dialectes arméniens*. Paris: Champion.
- AHLQVIST, ANDERS. 1988. Remarks on the question of dialects in Old Irish. *Historical dialectology: Regional and social*, ed. by Jacek Fisiak, 23–38. Berlin: Mouton de Gruyter.
- ALEKSEYENKO, ALEXANDER V.; CHRISTOPHER J. LEE; and MARC A. SUCHARD. 2008. Wagner and Dollo: A stochastic duet by composing two parsimonious solos. *Systematic Biology* 57.772–84.
- ANTHONY, DAVID W. 2007. *The horse, the wheel and language: How Bronze-Age riders from the Eurasian steppes shaped the modern world*. Princeton, NJ: Princeton University Press.
- ANTHONY, DAVID W. 2009. The Sintashta genesis: The roles of climate change, warfare, and long-distance trade. *Social complexity in prehistoric Eurasia*, ed. by Bryan K. Hanks and Katheryn M. Linduff, 47–73. Cambridge: Cambridge University Press.
- ANTHONY, DAVID W. 2013. Two IE phylogenies, three PIE migrations, and four kinds of steppe pastoralism. *Journal of Language Relationship* 9.1–21.
- ANTHONY, DAVID W., and DON RINGE. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1.199–219.
- ATKINSON, QUENTIN D., and RUSSELL D. GRAY. 2006. How old is the Indo-European language family? Illumination or more moths to the flame? In Forster & Renfrew, 91–109.
- ATKINSON, QUENTIN; GEOFF NICHOLS; DAVID WELCH; and RUSSELL GRAY. 2005. From words to dates: Water into wine, mathemagic or phylogenetic inference? *Transactions of the Philological Society* 103.193–219.
- BABEL, MOLLY; ANDREW GARRETT; MICHAEL HOUSER; and MAZIAR TOOSARVANDANI. 2013. Descent and diffusion in language diversification: A study of Western Numic dialectology. *International Journal of American Linguistics* 79.445–89.
- BAMMESBERGER, ALFRED. 1990. *Die Morphologie des urgermanischen Nomens*. Heidelberg: Carl Winter.

- BARBER, ELISABETH. 2001. The clues in the clothes: Some independent evidence for the movements of families. In Drews, 1–14.
- BEEKES, ROBERT S. P. 2010. *Etymological dictionary of Greek*. Leiden: Brill.
- BEEKES, ROBERT S. P. 2011. *Comparative Indo-European linguistics: An introduction*. 2nd edn. Amsterdam: John Benjamins.
- BELLWOOD, PETER. 2004. *First farmers: The origins of agricultural societies*. Malden, MA: Blackwell.
- BELLWOOD, PETER, and COLIN RENFREW (eds.) 2002. *Examining the farming/language dispersal hypothesis*. Cambridge: McDonald Institute for Archaeological Research.
- BERGSLAND, KNUD, and HANS VOGT. 1962. On the validity of glottochronology. *Current Anthropology* 3.115–53.
- BIRNBAUM, HENRIK. 1998. Common Slavic in time and space. *Scando-Slavica* 44.131–43.
- BIRNBAUM, HENRIK, and JAAN PUHVEL (eds.) 1966. *Ancient Indo-European dialects: Proceedings of the Conference on Indo-European Linguistics held at the University of California, Los Angeles, April 25–27, 1963*. Berkeley: University of California Press.
- BLUST, ROBERT. 2000. Why lexicostatistics doesn't work: The 'universal' constant hypothesis and the Austronesian languages. In Renfrew et al., vol. 2, 311–31.
- BLUST, ROBERT. 2004. *t to k: An Austronesian sound change revisited. *Oceanic Linguistics* 43.365–410.
- BOCQUET-APPEL, JEAN-PIERRE; STEPHAN NAJI; MARC VANDER LINDEN; and JANUSZ K. KOZŁOWSKI. 2009. Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of 14C dates. *Journal of Archaeological Science* 36. 807–20.
- 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* 337.957–60.
- 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. 2013. Corrections and clarifications. *Science* 342.1446.
- BOWERN, CLAIRE, and QUENTIN ATKINSON. 2012. Computational phylogenetics and the internal structure of Pama-Nyungan. *Language* 88.817–45.
- BOWERN, CLAIRE; PATIENCE EPPS; RUSSELL GRAY; JANE HILL; KEITH HUNLEY; PATRICK MCCONVELL; and JASON ZENTZ. 2011. Does lateral transmission obscure inheritance in hunter-gatherer languages? *PLoS ONE* 6.9.e25195. Online: <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0025195>.
- BRANDT, GUIDO; WOLFGANG HAAK; CHRISTINA J. ADLER; CHRISTINA ROTH; ANNA SZÉCSÉNYI-NAGY; SARAH KARIMNIA; SABINE MÖLLER-RIEKER; HARALD MELLER; ROBERT GANSLMEIER; SUSANNE FRIEDERICH; VEIT DRESELY; NICOLE NICKLISCH; JOSEPH K. PICKRELL; FRANK SIROCKO; DAVID REICH; ALAN COOPER; KURT W. ALT; and THE GENOGRAPHIC CONSORTIUM. 2013. Ancient DNA reveals key stages in the formation of Central European mitochondrial genetic diversity. *Science* 342.257–61.
- BUCK, CARL DARLING. 1949. *A dictionary of selected synonyms in the principal Indo-European languages*. Chicago: University of Chicago Press.
- BUCK, CARL DARLING. 1955. *The Greek dialects*. 2nd edn. Chicago: University of Chicago Press.
- CAMERON, ANGUS; ASHLEY CRANDELL AMOS; and ANTONETTE DI PAOLO HEALEY (eds.) 2007. *Dictionary of Old English: A to G online*. Toronto: Dictionary of Old English Project.
- CAMPBELL, LYLE, and WILLIAM J. POSER. 2008. *Language classification: History and method*. Cambridge: Cambridge University Press.
- CARDONA, GEORGE, and DHANESH K. JAIN (eds.) 2003. *The Indo-Aryan languages*. London: Routledge.
- CHANTRAINE, PIERRE. 1999. *Dictionnaire étymologique de la langue grecque*. 2nd edn. Paris: Klincksieck.
- CLACKSON, JAMES. 2000. Time depth in Indo-European. In Renfrew et al., vol. 2, 441–54.
- CLACKSON, JAMES. 2007. *Indo-European linguistics: An introduction*. Cambridge: Cambridge University Press.

- CLACKSON, JAMES, and GEOFFREY HORROCKS. 2011. *The Blackwell history of the Latin language*. Chichester: Wiley-Blackwell.
- COLVIN, STEPHEN. 2007. *A historical Greek reader: Mycenaean to the Koine*. Oxford: Oxford University Press.
- DARDEN, BILL. 2001. On the question of the Anatolian origin of Indo-Hittite. In Drews, 184–228.
- DAVIDS, T. W. RHYS, and WILLIAM STEDE. 1921–1925. *The Pali Text Society's Pali-English dictionary*. London: Luzac.
- DIAMOND, JARED, and PETER BELLWOOD. 2003. Farmers and their languages: The first expansions. *Science* 300.597–603.
- DREWS, ROBERT (ed.) 2001. *Greater Anatolia and the Indo-Hittite language family*. Washington, DC: Institute for the Study of Man.
- DRUMMOND, ALEXEI J.; SIMON Y. W. HO; MATTHEW J. PHILLIPS; and ANDREW RAMBAUT. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4.5.e88. Online: <http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0040088>.
- DRUMMOND, ALEXEI J.; MARC A. SUCHARD; D. XIE; and ANDREW RAMBAUT. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29.1969–73.
- DYEN, ISIDORE; A. T. JAMES; and J. W. L. COLE. 1967. Language divergence and estimated word retention rate. *Language* 43.150–71.
- DYEN, ISIDORE; JOSEPH B. KRUSKAL; and PAUL BLACK. 1992. *An Indo-European classification: A lexicostatistical experiment*. Philadelphia: American Philosophical Society.
- EMENEAU, MURRAY B. 1966. Dialects of Old Indo-Aryan. In Birnbaum & Puhvel, 123–38.
- ERNOUT, ALFRED, and ANTOINE MEILLET. 1951. *Dictionnaire étymologique de la langue latine: Histoire des mots*. 3rd edn. Paris: Klincksieck.
- EVANS, STEVEN N.; DONALD RINGE; and TANDY WARNOW. 2006. Inference of divergence times as a statistical inverse problem. In Forster & Renfrew, 119–29.
- FELSENSTEIN, JOSEPH. 1992. Phylogenies from restriction sites: A maximum-likelihood approach. *Evolution* 46.159–73.
- FELSENSTEIN, JOSEPH. 2004. *Inferring phylogenies*. Sunderland, MA: Sinauer.
- FORSTER, PETER, and COLIN RENFREW (eds.) 2006. *Phylogenetic methods and the prehistory of languages*. Cambridge: McDonald Institute for Archaeological Research.
- FORTSON, BENJAMIN W. 2010. *Indo-European language and culture: An introduction*. 2nd edn. Malden, MA: Blackwell.
- GAMKRELIDZE, THOMAS V., and VJAČESLAV V. IVANOV. 1995. *Indo-European and the Indo-Europeans*. Trans. by Johanna Nichols. Berlin: Mouton de Gruyter.
- GARRETT, ANDREW. 2006. Convergence in the formation of Indo-European subgroups: Phylogeny and chronology. In Forster & Renfrew, 138–51.
- GERAGHTY, PAUL. 1983. *The history of the Fijian languages*. Honolulu: University of Hawaii Press.
- GIMBUTAS, MARIJA. 1973. Old Europe c. 7000–3500 BC: The earliest European civilization before the infiltration of the Indo-European peoples. *Journal of Indo-European Studies* 1.1–21.
- GIMBUTAS, MARIJA. 1997. *The Kurgan culture and the Indo-Europeanization of Europe: Selected articles from 1952 to 1993*. Washington, DC: Institute for the Study of Man.
- GLARE, P. G. W. (ed.) 1982. *Oxford Latin dictionary*. Oxford: Oxford University Press.
- GRASSMANN, HERMANN. 1873. *Wörterbuch zum Rig-Veda*. Leipzig: F. A. Brockhaus.
- GRAY, RUSSELL D., and QUENTIN D. ATKINSON. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426.435–39.
- GRAY, RUSSELL D.; QUENTIN D. ATKINSON; and SIMON J. GREENHILL. 2011. Language evolution and human history: What a difference a date makes. *Philosophical Transactions of the Royal Society of London, Series B* 366.1090–1100. Online: <http://rspb.royalsocietypublishing.org/content/royptb/366/1567/1090>.
- GRAY, RUSSELL D.; ALEXEI J. DRUMMOND; and SIMON J. GREENHILL. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323.479–83.
- GRAY, RUSSELL D., and FIONA M. JORDAN. 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature* 405.1052–55.

- GREENHILL, SIMON J.; THOMAS E. CURRIE; and RUSSELL D. GRAY. 2009. Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society B* 276. 2299–306. Online: <http://rspb.royalsocietypublishing.org/content/276/1665/2299>.
- GREENHILL, SIMON J.; ALEXEI J. DRUMMOND; and RUSSELL D. GRAY. 2010. How accurate and robust are the phylogenetic estimates of Austronesian language relationships? *PLoS ONE* 5.3.e9573. Online: <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0009573>.
- GREENHILL, SIMON J., and RUSSELL D. GRAY. 2005. Testing population dispersal hypotheses: Pacific settlement, phylogenetic trees and Austronesian languages. *The evolution of cultural diversity: A phylogenetic approach*, ed. by Ruth Mace, Clare J. Holden, and Stephen Shennan, 31–52. London: University College London Press.
- GREENHILL, SIMON J., and RUSSELL D. GRAY. 2009. Austronesian language phylogenies: Myths and misconceptions about Bayesian computational methods. *Austronesian historical linguistics and culture history: A festschrift for Robert Blust* (Pacific linguistics 601), ed. by Alexander Adelaar and Andrew Pawley, 375–97. Canberra: Australian National University.
- GRIERSON, GEORGE ABRAHAM. 1932. *A dictionary of the Kashmiri language*. Calcutta: Asiatic Society of Bengal.
- HAHMO, SIRKKA-LIISA; TETTE HOFSTRA; ANDRIES DIRK KYLSTRA; and OSMO NIKKILÄ. 1991–2012. *Lexikon der älteren germanischen Lehnwörter in den ostseefinnischen Sprachen*, vols. 1–3. Amsterdam: Rodopi.
- HÄKKINEN, JAAKKO. 2012. Early contacts between Uralic and Yukaghir. *Per Urales ad Orientem: Iter polyphonicum multilinguae* (Suomalais-Ugrilaisen Seuran toimituksia 264), ed. by Tiina Hyytiäinen, Lotta Jalava, Janne Saarikivi, and Erika Sandman, 91–101. Helsinki: Suomalais-Ugrilainen Seura.
- HALWACHS, DIETER W.; BARBARA SCHRAMMEL; and ASTRID RADER. 2007. ROMLEX: The lexical database of Romani varieties. *Grazer Linguistische Studien* 67/68.117–32.
- HAUGEN, EINAR. 1976. *The Scandinavian languages: An introduction to their history*. Cambridge, MA: Harvard University Press.
- HEINE, BERND, and TANIA KUTEVA. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- HERMAN, JÓZSEF. 2000. *Vulgar Latin*. Trans. by Roger Wright. University Park: Pennsylvania State University Press.
- HIEBERT, FRANK. 1994. *Origins of the Bronze Age Oasis civilizations of Central Asia*. Cambridge, MA: American School of Prehistoric Research.
- HOCK, HANS HENRICH. 1991. Dialects, diglossia, and diachronic phonology in early Indo-Aryan. *Studies in the historical phonology of Asian languages*, ed. by William G. Boltz and Michael C. Shapiro, 119–59. Amsterdam: John Benjamins.
- HOLMAN, ERIC W.; CECIL H. BROWN; SØREN WICHMANN; ANDRÉ MÜLLER; VIVEKA VELUPILLAI; HARALD HAMMARSTRÖM; SEBASTIAN SAUPPE; HAGEN JUNG; DIK BAKKER; PAMELA BROWN; OLEG BELYAEV; MATTHIAS URBAN; ROBERT MAILHAMMER; JOHANN-MATTIS LIST; and DMITRY EGOROV. 2011. Automated dating of the world's language families based on lexical similarity. *Current Anthropology* 52.841–75.
- HORROCKS, GEOFFREY. 1997. *Greek: A history of the language and its speakers*. London: Longmans.
- HÜBSCHMANN, HEINRICH. 1897. *Armenische Etymologie*. Leipzig: Breitkopf & Härtel.
- IVANOV, VYACHESLAV V. 2001. Southern Anatolian and Northern Anatolian as separate Indo-European dialects and Anatolian as a late linguistic zone. In Drews, 131–83.
- JACKSON, KENNETH H. 1951. Common Gaelic: The evolution of the Goidelic languages. *Proceedings of the British Academy* 37.71–97.
- JANHUNEN, JUHA. 2009. Proto-Uralic—What, where, and when? *The quasiquicentennial of the Finno-Ugrian Society*, ed. by Jussi Ylikoski, 57–78. Helsinki: Société Finno-Ougrienne.
- JASANOFF, JAY H. 1994. Germanic (le germanique). *Langues indo-européennes*, ed. by Françoise Bader, 251–80. Paris: CNRS Editions.
- JASANOFF, JAY H. 2003. *Hittite and the Indo-European verb*. Oxford: Oxford University Press.
- JEFFREYS, HAROLD. 1961. *The theory of probability*. 3rd edn. Oxford: Oxford University Press.

- JENSEN, JOHAN LUDWIG WILLIAM VALDEMAR. 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica* 30.175–93.
- JOKI, AULIS J. 1973. *Uralier und Indogermanen: Die älteren Berührungen zwischen den uralischen und indogermanischen Sprachen*. Helsinki: Suomalais-Ugrilainen Seura.
- KALLIO, PETRI. 2005. A Uralic substrate in Balto-Slavic revisited. *Sprachkontakt und Sprachwandel*, ed. by Gerhard Meiser and Olav Hackstein, 275–83. Wiesbaden: Reichert.
- KALLIO, PETRI. 2006. On the earliest Slavic loanwords in Finnic. *The Slavization of the Russian north: Mechanisms and chronology*, ed. by Juhani Nuorluoto, 154–66. Helsinki: Helsinki University Press.
- KALLIO, PETRI. 2008. On the ‘early Baltic’ loanwords in Common Finnic. *Evidence and counterevidence: Essays in honour of Frederik Kortlandt*, ed. by Alexander Lubotsky, Jos Schaecken, and Jeroen Wiedenhof, vol. 1, 265–77. Amsterdam: Rodopi.
- KALLIO, PETRI. 2009. Stratigraphy of Indo-European loanwords in Saami. *Máttut – máddagat: The roots of Saami ethnicities, societies and spaces/places*, ed. by Tiina Äikäs, 30–45. Oulu: Giellagas Institute.
- KHANLARI, PARVIZ NATIL. 1979. *A history of the Persian language*. Trans. by N. H. Ansari. New Delhi: Sterling.
- KINGMAN, JOHN F. C. 1982. On the genealogy of large populations. *Journal of Applied Probability* 19.27–43.
- KOCH, JOHN T. (ed.) 2006. *Celtic culture: A historical encyclopedia*. Santa Barbara: ABC-CLIO.
- KOHL, PHILIP L. 2009. *The making of Bronze Age Eurasia*. Cambridge: Cambridge University Press.
- KOIVULEHTO, JORMA. 2001. The earliest contacts between Indo-European and Uralic speakers in the light of lexical loans. *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations*, ed. by Christian Carpelan, Asko Parpola, and Petteri Koskikallio, 235–63. Helsinki: Suomalais-Ugrilainen Seura.
- KRELL, KATHRIN S. 1998. Gimbutas’ Kurgan-PIE homeland hypothesis: A linguistic critique. *Archaeology and language II: Archaeological data and linguistic hypotheses*, ed. by Roger Blench and Matthew Spriggs, 267–82. London: Routledge.
- KURATH, HANS, and SHERMAN M. KUHN (eds.) 1952–2001. *Middle English dictionary*. Ann Arbor: University of Michigan Press.
- KUZ’MINA, ELENA E. 2007. *The origin of the Indo-Iranians*. Leiden: Brill.
- LAZARIDIS, IOSIF, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513.409–13.
- LEES, ROBERT B. 1953. The basis of glottochronology. *Language* 29.113–27.
- LEHRMAN, ALEXANDER. 2001. Reconstructing Proto-Indo-Hittite. In Drews, 106–30.
- LODGE, GONZALEZ. 1924–1933. *Lexicon Plautinum*. Leipzig: Teubner.
- LUBOTSKY, ALEXANDER. 2001. The Indo-Iranian substratum. *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations*, ed. by Christian Carpelan, Asko Parpola, and Petteri Koskikallio, 301–17. Helsinki: Société Finno-Ougrienne.
- LUICK, KARL. 1921–1940. *Historische Grammatik der englischen Sprache*. Leipzig: Tauchnitz.
- MALLORY, J. P. 1989. *In search of the Indo-Europeans: Language, archaeology and myth*. London: Thames & Hudson.
- MALLORY, J. P. 2002. Archaeological models and Asian Indo-Europeans. *Indo-Iranian languages and peoples*, ed. by Nicholas Sims-Williams, 19–42. Oxford: Oxford University Press.
- MALLORY, J. P., and DOUGLAS Q. ADAMS (eds.) 1997. *Encyclopedia of Indo-European culture*. London: Fitzroy Dearborn.
- MALLORY, J. P., and DOUGLAS Q. ADAMS. 2006. *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world*. Oxford: Oxford University Press.
- MASICA, COLIN P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- MATASOVIĆ, RANKO. 2009. *Etymological dictionary of Proto-Celtic*. Leiden: Brill.
- MAYRHOFER, MANFRED. 1989–2001. *Etymologisches Wörterbuch des Altindoarischen*. Heidelberg: Carl Winter.
- MCMAHON, APRIL, and ROBERT MCMAHON. 2006. Why linguists don’t do dates: Evidence from Indo-European and Australian languages. In Forster & Renfrew, 153–60.

- MEILLET, ANTOINE. 1920. *Aperçu d'une histoire de la langue grecque*. Paris: Klincksieck.
- MEILLET, ANTOINE. 1922. *Les dialectes indo-européens*. Paris: Champion.
- MEILLET, ANTOINE. 1925. *La méthode comparative en linguistique historique*. Paris: Champion.
- MELCHERT, H. CRAIG. 1998. The dialectal position of Anatolian within Indo-European. *Berkeley Linguistics Society (Special session on Indo-European subgrouping and internal relations)* 24.2.24–31. Online: <http://elanguage.net/journals/bls/article/view/3284>.
- MELCHERT, H. CRAIG. 2001. Critical response to the last four papers. In Drews, 229–35.
- 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* 70.545–66.
- NORMAN, KENNETH R. 1995. Dialect variation in Old and Middle Indo-Aryan. *The Indo-Aryans of ancient South Asia: Language, material culture and ethnicity*, ed. by George Erdosy, 278–92. Berlin: De Gruyter.
- OBERLIES, THOMAS. 1999. Middle Indo-Aryan and (the) Vedic (dialects). *Historische Sprachforschung* 112.39–57.
- OBERLIES, THOMAS. 2001. *Pāli: A grammar of the language of the Theravāda Tipiṭaka*. Berlin: Walter de Gruyter.
- OETTINGER, NORBERT. 1986. *'Indo-Hittite'-Hypothese und Wortbildung*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- OREL, VLADIMIR. 1998. *Albanian etymological dictionary*. Leiden: Brill.
- PAGEL, MARK; QUENTIN D. ATKINSON; and ANDREW MEADE. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449.717–20.
- PARPOLA, ASKO. 2008. Proto-Indo-European speakers of the late Tripolye Culture as the inventors of wheeled vehicles: Linguistic and archaeological considerations of the PIE homeland problem. *Proceedings of the nineteenth annual UCLA Indo-European Conference*, ed. by Karlene Jones-Bley, Martin E. Huld, Angela Della Volpe, and Miriam Robbins Dexter, 1–59. Washington, DC: Institute for the Study of Man.
- PARPOLA, ASKO. 2012. The problem of Samoyed origins in the light of archaeology: On the formation and dispersal of East Uralic (Proto-Ugro-Samoyed). *Per Urales ad Orientem: Iter polyphonicum multilinguae* (Suomalais-Ugrilaisen Seuran toimituksia 264), ed. by Tiina Hyytiäinen, Lotta Jalava, Janne Saarikivi, and Erika Sandman, 287–98. Helsinki: Suomalais-Ugrilainen Seura.
- PENZL, HERBERT. 1988. The Horn of Gallehus and the subgrouping of the Germanic languages. *Languages and cultures: Studies in honor of Edgar C. Polomé*, ed. by Mohammad Ali Jazayery and Werner Winter, 489–508. Berlin: Walter de Gruyter.
- PERLÈS, CATHERINE. 2001. *The early Neolithic in Greece: The first farming communities in Europe*. Cambridge: Cambridge University Press.
- PORZIG, WALTER. 1954. *Die Gliederung des indogermanischen Sprachgebiets*. Heidelberg: Winter.
- RÉDEI, KAROLY. 1986. *Zu den indogermanisch-uralischen Sprachkontakten*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- RENFREW, COLIN. 1987. *Archaeology and language: The puzzle of Indo-European origins*. London: Cape.
- RENFREW, COLIN. 1999. Time depth, convergence theory and innovation in Proto-Indo-European. *Journal of Indo-European Studies* 27.257–93.
- RENFREW, COLIN. 2000a. 10,000 or 5000 years ago?—Questions of time depth. In Renfrew et al., vol. 2, 413–39.
- RENFREW, COLIN. 2000b. At the edge of knowability: Towards a prehistory of languages. *Cambridge Archaeological Journal* 10.7–34.
- RENFREW, COLIN. 2001. The Anatolian origins of Proto-Indo-European and the autochthony of the Hittites. In Drews, 36–63.
- RENFREW, COLIN. 2003. Time depth, convergence theory and innovation in Proto-Indo-European: 'Old Europe' as a PIE linguistic area. *Languages in prehistoric Europe*, ed. by Alfred Bammesberger and Theo Vennemann, 17–48. Heidelberg: Winter.
- RENFREW, COLIN; APRIL MCMAHON; and LARRY TRASK (eds.) 2000. *Time depth in historical linguistics*. Cambridge: McDonald Institute for Archaeological Research.

- RIEKEN, ELISABETH. 2009. Der Archaismus des Hethitischen—eine Bestandsaufnahme. *Incontri Linguistici* 32.37–52.
- RINGE, DON. 2006. *From Proto-Indo-European to Proto-Germanic*. Oxford: Oxford University Press.
- RINGE, DON, and JOSEPH F. ESKA. 2013. *Historical linguistics: Toward a twenty-first century reintegration*. Cambridge: Cambridge University Press.
- RINGE, DON; TANDY WARNO; and ANN TAYLOR. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100.59–129.
- RIX, HELMUT; MARTIN KÜMMEL; THOMAS ZEHNDER; REINER LIPP; and BRIGITTE SCHIRMER (eds.) 2001. *LIV, Lexikon der indogermanischen Verben: Die Wurzeln und ihre Primärstambildungen*. 2nd edn. Wiesbaden: Ludwig Reichert.
- ROSS, MALCOLM. 1988. *Proto Oceanic and the Austronesian languages of western Melanesia*. Canberra: Australian National University.
- 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* 60.71–92.
- SALMONS, JOSEPH C. 2012. *A history of German: What the past reveals about today's language*. Oxford: Oxford University Press.
- SAPIR, EDWARD. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace.
- SENN, ALFRED. 1966. The relationships of Baltic and Slavic. In Birnbaum & Puhvel, 139–51.
- STADLER, TANJA. 2010. Sampling-through-time in birth-death trees. *Journal of Theoretical Biology* 267.396–404.
- STRIMMER, KORBINIAN, and OLIVER G. PYBUS. 2001. Exploring the demographic history of DNA sequences using the generalized skyline plot. *Molecular Biology and Evolution* 18.2298–305.
- STRUNK, KLAUS. 1994. Relative chronology and Indo-European verb-system: The case of the present- and aorist-stems. *Journal of Indo-European Studies* 22.417–34.
- STURTEVANT, EDGAR H. 1929. The relationship of Hittite to Indo-European. *Transactions and Proceedings of the American Philological Association* 60.25–37.
- STURTEVANT, EDGAR H. 1933. Archaism in Hittite. *Language* 9.1–11.
- SWADESH, MORRIS. 1952. Lexicostatistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96.452–63.
- SWADESH, MORRIS. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21.121–37.
- TAYLOR, ANN; TANDY WARNO; and DON RINGE. 2000. Character-based reconstruction of a linguistic cladogram. *Historical linguistics 1995*, ed. by John Charles Smith and Delia Bentley, vol. 1, 393–408. Amsterdam: John Benjamins.
- THOMASON, SARAH GREY. 2001. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.
- TIMBERLAKE, ALAN. 2013. Culture and the spread of Slavic. *Language typology and historical contingency*, ed. by Balthasar Bickel, Lenore A. Grenoble, David A. Peterson, and Alan Timberlake, 331–56. Amsterdam: John Benjamins.
- TISCHLER, JOHANN. 1973. *Glottochronologie und Lexikostatistik*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- TRAUGOTT, ELIZABETH CLOSS, and RICHARD B. DASHER. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.
- TURNER, RALPH L. 1962–1966. *A comparative dictionary of the Indo-Aryan languages*. London: Oxford University Press.
- URBAN, MATTHIAS. 2014. Lexical semantic change and semantic reconstruction. *The Routledge handbook of historical linguistics*, ed. by Claire Bowerman and Bethwyn Evans, 374–92. New York: Routledge.
- VÄÄNÄNEN, VEIKKO. 1981. *Introduction au latin vulgaire*. 3rd edn. Paris: Klincksieck.
- VAN ANDEL, TJEERD H., and CURTIS N. RUNNELS. 1995. The earliest farmers in Europe. *Antiquity* 69.481–500.
- VAUX, BERT. 2006. Armenian. *Encyclopedia of linguistics*, ed. by Philipp Strazny, 474–78. London: Routledge.

- WEISS, MICHAEL. 2014. The comparative method. *The Routledge handbook of historical linguistics*, ed. by Claire Bower and Bethwyn Evans, 127–45. New York: Routledge.
- WHEELER, WARD C., and PETER M. WHITELEY. 2014. Historical linguistics as a sequence optimization problem: The evolution and biogeography of Uto-Aztecan languages. *Cladistics*, early view. doi: 10.1111/cla.12078.
- WIENS, JOHN J., and DANIEL S. MOEN. 2008. Missing data and the accuracy of Bayesian phylogenetics. *Journal of Systematics and Evolution* 46.307–14.
- WILKINS, DAVID P. 1996. Natural tendencies of semantic change and the search for cognates. *The comparative method reviewed*, ed. by Mark Durie and Malcolm Ross, 264–304. Oxford: Oxford University Press.
- WINDFUHR, GERNOT, and JOHN J. PERRY. 2009. Persian and Tajik. *The Iranian languages*, ed. by Gernot Windfuhr, 416–44. London: Routledge.
- WITZEL, MICHAEL. 1999. Tracing the Vedic dialects. *Dialectes dans les littératures indo-aryennes*, ed. by Colette Caillat, 97–264. Paris: Collège de France, Institut de Civilisation Indienne.
- WITZEL, MICHAEL. 2003. *Linguistic evidence for cultural exchange in prehistoric western central Asia*. (Sino-Platonic papers 129.) Philadelphia: Department of East Asian Languages and Civilizations, University of Pennsylvania.
- WRIGHT, ROGER. 2011. Romance languages as a source for spoken Latin. *A companion to the Latin language*, ed. by James Clackson, 60–79. Malden, MA: Blackwell.
- YAKUBOVICH, ILYA. 2010. *Sociolinguistics of the Luvian language*. Leiden: Brill.
- YANG, ZIHENG. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *Journal of Molecular Evolution* 39.306–14.

[wchang@gmail.com] (Chang)
[chundra@berkeley.edu] (Cathcart)
[david.lw.hall@gmail.com] (Hall)
[garrett@berkeley.edu] (Garrett)

[Received 10 June 2014;
accepted 22 September 2014]