Transactions of the Philological Society Volume 103:2 (2005) 193–219

FROM WORDS TO DATES: WATER INTO WINE, MATHEMAGIC OR PHYLOGENETIC INFERENCE?

By Quentin Atkinson[a], Geoff Nicholls[b], David Welch[b] and Russell Gray[a]
[a]*Department of Psychology,* [b]*Department of Mathematics, University of Auckland*

Abstract

Gray & Atkinson's (2003) application of quantitative phylogenetic methods to Dyen, Kruskal & Black's (1992) Indo-European database produced controversial divergence time estimates. Here we test the robustness of these results using an alternative data set of ancient Indo-European languages. We employ two very different stochastic models of lexical evolution – Gray & Atkinson's (2003) finite-sites model and a stochastic-Dollo model of word evolution introduced by Nicholls & Gray (in press). Results of this analysis support the findings of Gray & Atkinson (2003). We also tested the ability of both methods to reconstruct phylogeny and divergence times accurately from synthetic data. The methods performed well under a range of scenarios, including widespread and localized borrowing.

1. Introduction

Questions about the origins of human populations hold an enduring fascination. Often the answer lies beyond recorded history, prior to even the oldest of ancient manuscripts or oral traditions. In such cases, although we may not have access to any literal description of events, we can nonetheless gather other kinds of evidence to determine what happened when. Thomas Jefferson, for example, in *Notes on the State of Virginia* (1782: 227), used an intuitive comparison of related languages to draw historical inferences about their age –

Published by Blackwell Publishing,
9600 Garsington Road, Oxford OX4 2DQ and 350 Main Street, Malden, MA 02148, USA.

> A separation into dialects may be the work of a few ages only, but for two dialects to recede from one another till they have lost all vestiges of their common origin, must require an immense course of time; perhaps not less than many people give to the age of the earth. *A greater number of those radical changes of language having taken place among the red men of America, proves them of greater antiquity than those of Asia.*

Jefferson's reasoning was based on the implicit assumption that languages change through a process of descent with modification and that this change occurs at a roughly constant rate. Debate over whether or not this assumption is plausible has generated as much controversy as the historical problems to which it has been applied. Can word lists be turned into dates, and can we quantify the uncertainty in these time estimates? We will argue here that, at least in the case of Indo-European, the answer to both of these questions is "yes".

2. Words into Dates or Water into Wine?

Some 170 years after Jefferson, Morris Swadesh (1952, 1955) formalized the idea of inferring language divergence times from word lists when he developed lexicostatistics and glottochronology. Lexicostatistical methods infer language trees from the percentage of shared cognates between languages – the more similar the languages, the more closely they are related. Usually a list of 100 or 200 meanings (known as the Swadesh 100 and 200 word lists) are used, comprising terms that are thought to be relatively universal, culture-free and resistant to borrowing. Glottochronology is an extension of this approach to estimate time depth on the resulting genealogical tree under the assumption of a 'glotto-clock' or constant rate of lexical evolution. Using traditional glottochronology, time-depth is calculated using the formula $t = (\log C)/(2 \log r)$, where t is time depth in millennia, C is the percentage of cognates shared and r is the "universal" constant or rate of retention (the expected proportion of cognates remaining after 1,000 years of separation; Swadesh, 1955). For the Swadesh 200 word list, a value of 81% is often used for r.

Although initially received with some enthusiasm, glottochronology was heavily criticised and most linguists today view it as discredited. These well known criticisms are discussed in some detail in Campbell (2004), Gray & Atkinson (2003) and Atkinson & Gray (in press[a, b]). There are four key criticisms of glottochronology. First, the conversion of character-state data to percentage similarity scores between languages results in a loss of information and hence a reduction in the power of the method to reconstruct tree topology and branch lengths accurately (Steel, Hendy & Penny, 1988). This problem is exacerbated by the fact that some languages have multiple words for a given meaning and a few may have no word or the same word for two different meanings (Campbell, 2004). Second, distance-based tree-building techniques such as UPGMA (Unweighted Pair Group Method with Arithmetic mean) can produce erroneous trees under conditions of rate heterogeneity (Blust, 2000). Third, language evolution is not always tree-like. Terms can be borrowed, distorting divergence time estimates. Fourth, by comparing ages estimated using glottochronology with historically attested dates, Bergsland & Vogt (1962) demonstrated that glottochronology can produce erroneous results due to rate variation.

Gray & Atkinson (2003; Atkinson & Gray, in press[a, b]) argue that whilst the criticisms of glottochronology reflect legitimate concerns, it is possible to estimate dates using a different class of methods. Current statistical phylogenetic methods used widely in biology allow us to overcome the problems associated with glottochronology. First, character-based stochastic models of evolution retain phylogenetic information from the source data and allow us to reconstruct phylogeny accurately, even under conditions of rate heterogeneity (Huelsenbeck *et al.*, 2001). Character-based methods can also account for polymorphisms (multiple words for the same meaning) and uncertainty in vocabulary assignments and cognacy judgements. Further, by using an explicit model of evolution, the assumptions of the model are clear and the effects of changing these assumptions can be tested easily. Second, Bayesian phylogenetic inference allows us to quantify random error in tree topology and branch-lengths, a crucial factor if results are to be used to test between competing hypotheses. Third, the degree of reticulate evolution and borrowing between languages can be assessed using visualization tools, like Split Decomposition (Huson, 1998) and NeighborNet (Bryant & Moulton, 2002), that do not assume a tree-like model of evolution (Bryant, Filimon & Gray, in press). A statistical framework also makes it possible to test the robustness of results to some violations in the assumptions of the method, including the nature and magnitude of borrowing. For example, Atkinson & Gray (in press[b]) presented evidence that their divergence time estimates were robust to borrowing by comparing a number of different coding procedures and analysing subsets of the lexicon thought to be more resistant to borrowing. Nicholls & Gray (in press) tested the robustness of their results to borrowing also, using synthetic data to simulate various degrees of reticulation. Finally, violations of the assumption of rate constancy can be investigated in a similar fashion. Where there is a concern about its effect, rate-smoothing algorithms allow divergence times to be estimated without the assumption of a strict glotto-clock. Gray & Atkinson (2003) used this approach to test between competing hypotheses for the age of Indo-European. Nicholls & Gray (in press) quantify the uncertainty introduced by rate variation by analysing subsets of the data for which historically attested dates exist.

Many features of vocabulary evolution are not represented in the classes of models that Gray & Atkinson (2003) and Nicholls & Gray (in press) fit to lexical data. As a consequence uncertainty arises from two sources:*random error*, which would be present if their models were perfect descriptions of vocabulary evolution, and which arises from stochastic fluctuations predictable in distribution; and *systematic error*, the estimation bias caused by model misspecification. Random error will be well-estimated in the Bayesian framework employed here. However, the problem of quantifying the uncertainty due to model misspecification is much harder, since the number of “real” language evolution models one might reasonably entertain is enormous. Progress can be made only by focusing on those effects thought likely to be important and estimating the size (and direction) of the biases they cause (see section 8).

3. A New Set of Ancient Data

Gray & Atkinson (2003) based their analysis on lexical data derived from Dyen, Kruskal & Black's (1992) Indo-European database. They used a modified electronic version of the Dyen, Kruskal & Black (1997) dataset comprising 200 Swadesh list meanings in 87 languages with cognacy judgements made by expert linguists. Three extinct languages were added using multiple sources to corroborate cognacy judgements (Adams, 1999; Gamkrelidze & Ivanov, 1995; Guterbock & Hoffner, 1986; Hoffner, 1967; Tischler, 1973, 1997). Divergence time estimates from Gray & Atkinson (2003) suggested a root age of Indo-European of between 7,800 and 9,800 BP, consistent with the Anatolian theory of Indo-European origin (Renfrew, 1987). Crucially, this age range was outside the 5,000 to 6,000 BP age range implied by the alternative, Kurgan, theory of Indo-European origin (Gimbutas, 1973a,b). These results were supported by further work from Atkinson & Gray (in press[a, b]) and Nicholls & Gray (in press).

Here, we apply a number of new techniques and tools of analysis to an alternative Indo-European dataset compiled by Ringe, Warnow & Taylor (2002). This data includes 430 meaning categories in 20 extinct and 4 extant Indo-European languages. Table 1 shows a sub-sample of the data with eight meaning categories in four Germanic languages and Greek.

Repeating the two analyses of Gray & Atkinson (2003) and Nicholls & Gray (in press) on a second Indo-European dataset has a number of benefits. The data include different languages and some

Table 1. A sample dataset of eight meaning categories across four Germanic languages and Greek. Numbers in superscript indicate cognate sets within each meaning category identified by Ringe, Warnow and Taylor (2002).

Meaning	all	fall	I	leg	mountain	pull	sing	water
Old English	ealle^{1}	fielþ^{1}	ic^{1}	scanca^{1}	beorg^{1}	tīehþ^{1}, drægþ^{3}	singþ^{1}	wæter^{1}
Old High German	alle^{1}	fellit^{1}	ih^{1}	bein^{2}	berg^{1}	dinsit^{2}, ziuhit^{1}	singit^{1}	wazzar^{1}
Old Norse	allir^{1}	fellr^{1}	ek^{1}	leggr^{3}	fjall^{2}	dregr^{3}	syngr^{1}	vatn^{1}
Gothic	allai^{1}	driusiþ^{2}	ik^{1}	–	faírguni^{3}	atþinsiþ^{2}	siggwiþ^{1}	watō^{1}
Greek	*πάυτες*^{2}	*πίπτει*^{3}	*ἐγώ*^{1}	*σκέλος*^{4}	*ὅρος*^{4}	*ἕλκει*^{4}	*ἄιδει*^{2}	*ὕδωρ*^{1}

different meaning categories to those in the Dyen, Kruskal & Black (1992) database. In addition, cognacy judgements were made by Don Ringe, an independent expert in the field, who was not involved with the Dyen, Kruskal & Black (1992) study. These facts alone make this analysis a particularly interesting test of the methodology. Different date estimates for each set of data would cast doubt on the reliability of the method and/or the amount of temporal information in the data. Conversely, consistent dates across both datasets would support the idea that there is a strong temporal signal in lexical data and that the methods employed are robust to variations in the cognacy judgement criteria, meaning categories used and even the languages analysed.

Of course, agreement across data sets might simply indicate a bias in the inferential methodology. If this were the case we would expect the methodology to show similar biases on some synthetic data. In section 8, we test the methodology on a range of synthetic data simulated under models that were designed to include features thought to be problematic for the methodology. In this way we deliberately set up model misspecification and test for its effect.

Another advantage of using the Ringe *et al.* (2002) data is that it comprises mainly ancient languages. First, all other things being equal, this should improve the resolution of some of the deeper Indo-European branching structure. Second, Garrett (in press) has found evidence for modern advergence processes in a number of Indo-European sub-groups. He points out that the analyses of Gray & Atkinson (2003; Atkinson & Gray, in press[a, b]) and Nicholls & Gray (in press) used mainly contemporary languages and hence may be biased by the effect of certain types of unidentified modern borrowing. Atkinson & Gray (in press[b]) discuss why this is unlikely and Nicholls & Gray (in press) argue that sampling fewer languages lessens the impact of borrowing. However, here, by analysing a dataset of ancient languages, many of which are two or three thousand years old, we can all but eliminate the effect of modern borrowing. Moreover, our synthetic studies show that if local borrowing is present in ancient and modern languages this need not produce a bias. Where model distortion is mild, and uniform over time, parameters distorted to fit calibration points near the leaves predict ages for unattested branching events.

4. ENCODING WORD LISTS FOR CHARACTER BASED ANALYSIS

Unlike glottochronology, we never count cognates nor calculate ages based on pair-wise distances between languages. Character-based methods ensure that information about the presence or absence of individual word forms or grammatical and phonological features (the characters) can be retained. Table 2 shows a matrix of 1's and 0's that expresses the presence or absence of cognates across the five languages listed in Table 1.

5. STOCHASTIC MODELS AND BAYESIAN INFERENCE OF PHYLOGENY

Stochastic models of evolution and Bayesian inference of phylogeny allow us to overcome the problems identified with the distance-based tree-building methods used in lexicostatistics and glottochronology. Bayesian and related likelihood-based inference can outperform distance and parsimony tree-building methods in situations where models are reliable and there are unequal rates of change (Kuhner & Felsenstein 1994). Likelihood-based inference integrates three related components – the observed data, a stochastic model of character evolution, and an evolutionary tree or a set of trees – and is based on the premise that we should favour the explanation that makes our observed data most likely. For languages, the observed data can take the form of a binary matrix

Table 2. Indo-European cognates from table 1 expressed in a binary matrix. Each row represents a language and each column a character (in this case a cognate set). A '1' indicates that a particular cognate set is present in that language, a '0' indicates absence of the cognate set and a '?' indicates uncertainty.

Meaning	all		fall			I	leg				mountain				pull				sing		water
Cognate set	1	2	1	2	3	1	1	2	3	4	1	2	3	4	1	2	3	4	1	2	1
Old English	1	0	1	0	0	1	1	0	0	0	1	0	0	0	1	0	1	0	1	0	1
Old High German	1	0	1	0	0	1	0	1	0	0	1	0	0	0	1	1	0	0	1	0	1
Old Norse	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0	1	0	1
Gothic	1	0	0	1	0	1	?	?	?	?	0	0	1	0	0	1	0	0	1	0	1
Greek	0	1	0	0	1	1	0	0	0	1	0	0	0	1	0	0	0	1	0	1	1

that codes cognate presence or absence, as in Table 2. It is assumed that these data are generated by some stochastic model of character evolution on the tree. This model is called the "observation model". Trees which make the sequence data a relatively more likely outcome of the observation model have relatively higher likelihood scores. The maximum likelihood (ML) tree is that tree or trees making the data most likely. Although parameters of the character evolution model can be assigned values prior to analysis, it is possible, and usually preferable, to estimate these observation-model parameters from the data. This allows *a priori* assumptions about the process of character evolution to be minimized and is a major advantage of the likelihood approach to phylogenetic inference (Pagel 1997).

The basic procedure for calculating the likelihood score for the model of Gray & Atkinson (2003) is described in Atkinson & Gray (in press[a]), and is provided as online supplementary material to this paper (available at http://ling.man.ac.uk/More/PhilSoc/Transactions.html) – a more detailed explanation in a biological context can be found in Swofford *et al.* (1996). It is natural to present the age of the Most Recent Common Ancestor (MRCA) of all languages in the ML tree as the "result" of an analysis. However, there are usually many trees with likelihood scores which are close to the likelihood of the ML tree. It is easy to show that even when the observation model is an accurate description of cognate evolution there is a very high probability that the true tree will not coincide with the ML tree. For this reason, where feasible, it is preferable to report a confidence interval for the age of the MRCA which takes into account uncertainty in the reconstructed tree.

To do this we used Bayesian inference of phylogeny and Markov Chain Monte Carlo (MCMC) algorithms (Metropolis *et al.*, 1953) to generate a sample distribution of trees that reflects the component of phylogenetic uncertainty in our analysis due to random error. This inference is based on the data, an observation model and a set of prior beliefs (or *priors*) about all unknown parameters of the model, including the tree topology, branch-lengths, and rate matrix. We favour priors that are uninformative with respect to the hypotheses being investigated. However, an important component of any Bayesian MCMC analysis is to ensure that results are robust to a range of sensible priors. A summary explanation of the application of Bayes' theorem and the MCMC algorithms to phylogenetics is included in the supplementary material, and discussed in more detail in Atkinson & Gray (in press[b]) and Huelsenbeck *et al.* (2001).

Crucially, the MCMC posterior sample of trees is much more informative about the phylogenetic signal in the data than methods that return any single "optimal" phylogeny (c.f. Ringe *et al.*, 2002). The sample distribution allows us to approximate phylogenetic uncertainty (uncertainty in tree topology and branch lengths), given the data, and to incorporate this into our results. This is impossible using either the comparative method or traditional glottochronology, and yet the effect of phylogenetic uncertainty is a crucial consideration if results are to be used to test historical hypotheses. We can calculate divergence times across the entire Bayesian sample distribution and determine the error in date estimates resulting from phylogenetic uncertainty. This also means that we can make inferences about the age of Indo-European without having to commit to a particular topology.

6. Two Different Approaches to Likelihood Inference and Modelling

Most scientists would agree that the best way to validate a result is to repeat the analysis, preferably on an independent data set. A close second, however, may be to reanalyse the data using a different methodology and model. Here then, as well as analysing an alternative dataset, we employ two very different models of language evolution, based on a very different set of core assumptions. Each model and methodology is described briefly below. A more detailed explanation is available in the online supplementary material (http://www.philsoc.org.uk), and in Atkinson & Gray (in press [a, b]) and Nicholls & Gray (in press).

6.1 Finite-sites Model and Method 1

Gray & Atkinson (2003; and Atkinson & Gray, in press[a, b]) use a model of binary character evolution implemented in the programme

MrBayes (Huelsenbeck & Ronquist, 2001) to generate a sample distribution of trees with branch lengths proportional to the inferred amount of evolutionary change. We can model the process of cognate gain (0 to 1) and loss (1 to 0) using three parameters – μ, the mean substitution rate, and π^0 and π^1, which represent the relative frequencies of 1's and 0's. The substitution rate is a parameter estimated in the MCMC analysis and the equilibrium frequency of 1's and 0's can be estimated from their frequency in the data. Missing data is treated as another unknown binary parameter to be estimated.

Because there are just two states, 0 and 1, this model is trivially time-reversible – we cannot tell the direction in which the cognate evolved from its history in a single language. This model allows a single cognate to appear in and disappear from a single language more than once over the course of time, allowing the model to mimic the effect of word-borrowing. As the direction of time is not determined, we cannot determine the root of the tree from the data – we need to provide an outgroup as a root. For all the method 1 analyses reported here, trees were rooted with Hittite, consistent with independent linguistic analyses (Gamkrelidze & Ivanov, 1995; Rexova, Frynta & Zrzavy, 2003). Atkinson & Gray (in press[a, b]) found that the root point did not affect age estimates significantly. A Gamma shape parameter (α) was also added to allow for rate variation between cognate sets. As with the overall rate parameter, α was estimated from the data. An α value of 5 was observed, indicating moderate rate variation between cognate sets.

We can also account for rate variation between lineages and through time by relaxing the assumption of a strict glottoclock. The 87 languages in the modified Dyen *et al.* (1997) data set allowed for 11 internal clade constraints (see supplementary material). Terminal nodes representing contemporary languages were set to 0 years whilst 3 extinct languages (Hittite and Tocharian A & B) were constrained in accordance with estimated ages of the source texts. For the 24 languages in the Ringe *et al.* (2002) data, 12 internal node constraints were available, whilst 20 extinct languages were constrained in accordance with estimated ages of the source texts (see supplementary material). Sanderson's (2002a) penalized-likelihood algorithm, as implemented in *r8s*

(Sanderson, 2002b), was then used to smooth rates of evolution across each tree and to calculate divergence times. This approach allows for rate variation between lineages while incorporating a "roughness penalty" that costs the model more if rates vary excessively from branch to branch. Interestingly, high smoothing factors were found to fit the data best, suggesting that the process of evolution is in fact relatively tightly constrained. The procedure was repeated on all of the trees in the MCMC Bayesian sample distribution. The distribution of divergence times at the root can be used to create a confidence interval for the age of Indo-European.

There are three potential criticisms of this model. First, the same rate parameter is used to estimate cognate gains and losses. Whilst cognates can be lost relatively easily, the innovation events that produce them are rare - it is very unlikely that two languages would ever independently gain the same cognate. Thus, it may be argued that trying to fit a single rate parameter to a model of cognate gain and loss is "patently inappropriate" (Evans, Ringe & Warnow, in press). Indeed, this may be problematic if the rate of gain and loss are widely different. However, processes of borrowing and dialect chains at divergence mean that models which allow cognates to be gained more than once may still be reasonable. In fact, as we will see below, this feature of the model may allow it to accommodate moderate reticulation in the data.

A second criticism is that the inference is itself a hybrid of ML and Bayesian inference. The distribution over unrooted trees is a posterior distribution, but the estimation of branch lengths is a penalized ML method. This will cause problems if, for example, there is significant uncertainty in branch lengths, so that the mode of the branch length distribution is not representative of the branch length distribution as a whole.

A third potential criticism is that method 1 uses a "finite sites" model from biology. This means character state changes are modelled through time across a fixed number of characters, or 'sites'. However, in reality, the number of cognates or sites is not finite and depends on the number of languages we are looking at and how long they have been evolving. In what follows we describe an alternative model introduced by Nicholls & Gray (in press) that does not assume finite sites.

6.2 Stochastic-Dollo Model and Method 2

Dollo's Law states that traits can evolve only once (Farris, 1977). In this context, we treat cognates as traits and assume that the same cognate cannot be independently created in different languages. This assumption is equivalent to asserting that the cognate data is homoplasy free (c.f. Ringe *et al.*, 2002). Based on this assumption, we outline a stochastic model of language change appropriate to the cognate data described in section 3.

The model allows language change to occur in three different ways: words (and corresponding cognate sets) are created, words are lost, and words reproduce (when languages split, forming two child copies of a parent language). We assume that words are created in any given language at rate λ. When a word is created, it falls into a new cognate class, so word creation and cognate class creation are synonymous. If there are k languages extant at time t, new cognates are created at total rate $k\lambda$. Each word is lost from a given language independently at rate μ. If at time t, there are k languages and language i contains l_i words, word death occurs at a total rate of $\mu(l_1 + l_2 + \ldots + l_k)$.

Each language splits at rate θ. When a language splits, two child copies of the language are made and the parent language dies. At the time of splitting, the child languages are indistinguishable from the parent language and thereafter evolve in exactly the same way as the parent language did. If there are k languages at time t, language splitting occurs at total rate $k\theta$.

We assume that the times between all events causing language change are exponentially distributed and that all rates – the cognate birth rate, λ, the cognate loss rate, μ, and the language splitting rate, θ – are constant across time and space. We assume also that all languages and cognates evolve independently.

The data described in section 2 is collected in such a way that cognates that are present in no languages or only one language at the time of collection are not recorded. Thus the observed cognate birth rate λ^* is different from the actual cognate birth rate λ since words must be born and survive into at least two languages in order to be observed. This data thinning process may result in the birth times of cognates in the data being unevenly distributed over the tree. This effect is accounted for in the likelihood calculation for a given tree, the details of which are given in Nicholls & Gray (in press).

There are two obvious features of the data that this model fails to capture. The first is missing data. We do not account for missing data and recode any missing cognates as absent. It is necessary to check for biases caused by this approximation. We repeat analyses omitting languages with a significant amount of missing data (we use 13% as the cut-off below). The effect of doing this on the relatively complete data sets treated in this paper is negligible.

A more important issue is that of borrowing between languages. If one language gains a new word by borrowing it from another, the Dollo assumption is violated. While it is relatively simple to include borrowing when building a model of language change, we are currently unable to analyse such a model. In order to quantify the magnitude of this misspecification, in section 8 we present a series of analyses of data synthesised under models with borrowing but analysed under the Stochastic-Dollo model.

Inference for the Stochastic-Dollo model is made within a Bayesian framework and the data is analysed using a MCMC algorithm implemented in Matlab by two of the authors (GN and DW). The relevant software, called TraitLab, can be downloaded from (aitken.math.auckland.ac.nz/~nicholls/TraitLab/).

6.3 Other inference issues

Both of the above models assume that the characters in the binary data are independent. Evans *et al.* (in press) argue that the results of any analysis using these models are invalid due to violations of the assumption of independence. They make the important point that the independence assumption is violated when individual meanings in the Swadesh word list are broken up into characters representing multiple cognate sets. Specifically, if a particular cognate set is present in a language, it will be less likely that other cognate sets for the same meaning will also be present. Conversely the core meanings must be occupied at all times in all languages, whereas both models allow core meaning categories to be empty. We attempt to detect bias caused by certain effects in this class below using synthetic data. The problem of empty meaning categories does not seem to be important. We surmise that this is because ancestral meaning categories need not be filled by cognates present in the data. They may be occupied by ancient cognates with no instances in the data.

In characterizing their data, Ringe *et al.* (2002) are concerned with the distinction between shared innovations and shared retentions. In both methods described here, we form a model weighted average over these cases and, as a consequence, this distinction is not a problem.

7. RESULTS

Figure 1 shows the results of a series of analyses of both data sets using both of the models of evolution described above. Table 3 summarizes these results including the data, priors and other conditions used for each analysis from Figure 1.

Gray & Atkinson (2003) found divergence time estimates for the root of the Indo-European tree were robust to a wide range of plausible rooting points, Bayesian priors, cognacy judgement criteria, age constraints and the effect of missing information in the data. Key results from Gray & Atkinson (2003) are summarized in figure 1 (DF1-6). These results are consistent with a number of subsequent analyses, using subsets of 20 languages (DF8 & 9) and even when the data is limited to the highly conserved and borrowing-resistant Swadesh 100 word list (DF7). Using their stochastic-Dollo model, Nicholls & Gray (in press) found evidence for similar ages using the Swadesh 100 word list items (DD11) and slightly younger divergence times using the whole data set (DD10) and a subset of 31 languages (DD12). The inferred ages are broadly consistent with the Gray & Atkinson (2003) results.

Results from the analysis of the Ringe *et al.* (2002) data are consistent with the Dyen *et al.* (1997) data. Varying branch-length priors (RF2), rates across sites (RF7), and rates through time (RF8), and the cognacy judgement criteria (RF3, RD10) had little

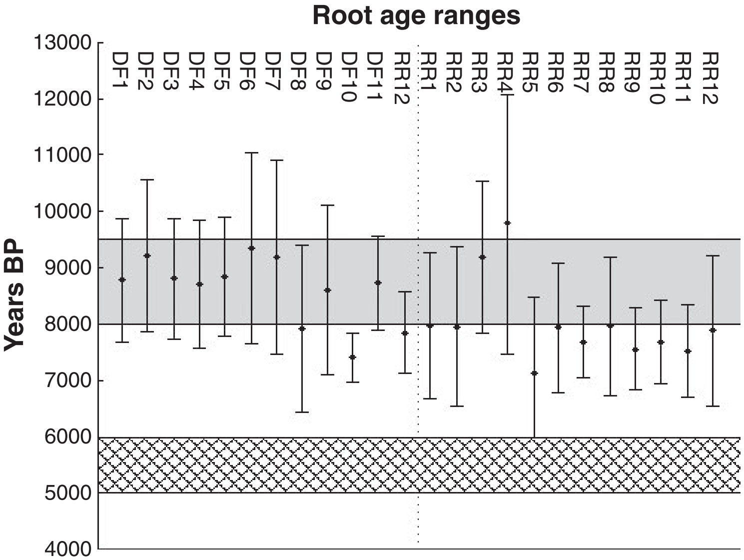


Figure 1. Results of analyses of the Dyen *et al.* (1997) dataset (to the left of the dotted line, with labels beginning with 'D') and Ringe *et al.* (2002) dataset (to the right of the dotted line, with labels beginning with 'R') using the finite-sites (labelled 'DF' and 'RF') and stochastic-Dollo (labelled 'DD' and 'RD') models. Each analysis is summarized with a plotted mean and error bars representing a 95% confidence interval. The horizontal bands indicate the age range implied under the two competing theories of Indo-European origin – the Kurgan hypothesis (cross-hatched band) and the Anatolian hypothesis (grey band).

effect on divergence times estimated from the Ringe *et al.* data. Analysing progressively more refined data sets (Swadesh 200 word list items only – RF4, RD11; and Swadesh 100 word list items only – RF5, RD12), also had little effect on the mean age estimates, although, predictably, as the amount of information decreased, variance in the estimated dates increased. The analysis RF5, of Swadesh 100 list terms, appears to go against this trend, showing less variation than the analysis of Swadesh 200 list terms in RF4.

Table 3. Summary of analyses from figure 1, including the mean and standard deviation for the age at the root of Indo-European.

#	Data	Model	μ	S.D	Comments
DF1	Dyen *et al.* (1997)	Finite-sites	8764	544	All cognacy information, uniform branch length priors, Gamma distributed rates across sites, PL rate-smoothing.
DF2	Dyen *et al.* (1997)	Finite-sites	9201	670	As for DF1 with conservative cognacy judgements only.
DF3	Dyen *et al.* (1997)	Finite-sites	8794	533	As for DF1 with loose constraints.
DF4	Dyen *et al.* (1997)	Finite-sites	8699	561	As for DF1 with missing data coded as '?'
DF5	Dyen *et al.* (1997)	Finite-sites	8829	520	As for DF3 with strict clock assumption.
DF6	Dyen *et al.* (1997)	Finite-sites	9336	843	As for DF2 with exponential branch length priors.
DF7	Dyen *et al.* (1997)	Finite-sites	9176	855	As for DF2 with Swadesh 100 word list terms only.
DF8	Dyen *et al.* (1997)	Finite-sites	7900	735	As for DF2 with selection of 20 languages.
DF9	Dyen *et al.* (1997)	Finite-sites	8590	748	As for DF8 with selection of a different 20 languages.
DD10	Dyen *et al.* (1997)	Dollo	7400	218	Conservative cognacy judgements, uniform MRCA-time prior.
DD11	Dyen *et al.* (1997)	Dollo	8714	410	As for DD10 with Swadesh 100 word list terms only.
DD12	Dyen *et al.* (1997)	Dollo	7834	360	As for DD10 with selection of 31 languages
RF1	Ringe *et al.* (2002)	Finite-sites	7964	645	Uniform branch length priors, gamma distributed rates across sites, PL rate-smoothing
RF2	Ringe *et al.* (2002)	Finite-sites	7942	702	As for RF1 with exponential branch length priors.
RF3	Ringe *et al.* (2002)	Finite-sites	9166	671	As for RF1 with fine grained cognacy judgements for some characters.
RF4	Ringe *et al.* (2002)	Finite-sites	9770	1150	As for RF1 with Swadesh 200 word list items only.
RF5	Ringe *et al.* (2002)	Finite-sites	7106	671	As for RF1 with Swadesh 100 word list items only.
RF6	Ringe *et al.* (2002)	Finite-sites	7932	572	As for RF1 with topology constraint in accordance with Ringe *et al.* (2002).
RF7	Ringe *et al.* (2002)	Finite-sites	7665	313	As for RF1 with equal rates across sites.
RF8	Ringe *et al.* (2002)	Finite-sites	7956	612	As for RF1 with strict clock
RD9	Ringe *et al.* (2002)	Dollo	7552	366	Uniform MRCA-time prior, 15 languages with over 10% sampling.
RD10	Ringe *et al.* (2002)	Dollo	7671	368	As for RD9 with fine grained cognacy judgements for some characters.
RD11	Ringe *et al.* (2002)	Dollo	7513	406	As for RD9 with Swadesh 200 word list items only.
RD12	Ringe *et al.* (2002)	Dollo	7869	659	As for RD9 with Swadesh 100 word list items only, 17 languages with over 10% sampling.

This is misleading, however, as the weak signal in the data meant that most of the trees produced in this analysis had to be filtered out because they were inconsistent with the clade constraints for known Indo-European language groups. As a result, variation was artificially decreased in this analysis, a case where Method 1 fails to produce reliable error bars. In analyses of larger data sets almost all of the trees in the Bayesian sample were consistent with recognized Indo-European groups. This anomaly did not occur using the Trait Lab analysis of the Swadesh 100 dataset because the programme allows topology to be constrained prior to the analysis. Finally, in RF6, the topology of the Indo-European tree was constrained to that obtained by Ringe *et al.*'s (2002) own analysis of lexical, phonological and morphological characters. Again, the estimated age at the root of the tree is unaffected. This is a consequence of the fact that data can be informative of total tree length even when uninformative of topology.

8. Controlled Miracles – Synthetic Data Validation

No statistical model of language change will capture all aspects of a process as complicated as the evolution of a language lexicon. In this sense all models are lies. However, models can be used as "lies that lead us towards the truth". In other words, they can allow us to answer questions of interest with sufficient precision and accuracy so as to provide a meaningful result. The crucial caveat is that we must test the performance of our model for a given task. To determine precision, we need to be able to quantify the uncertainty in our results. One way to do this is to generate synthetic data on a known phylogeny under plausible alternative models of evolution, including models which violate the assumptions of our inference model. Validating a methodology on synthetic data is useful for a number of reasons. First, we know the "true" phylogeny - the topology and branch lengths on which the data was generated. This means we can test the ability of our method to reconstruct a phylogeny from a given set of data accurately. Second, it is generally much easier to generate data on a phylogeny than it is to estimate the phylogeny that has produced a given set of data. As a result, we can test the performance of our methods on data generated under more sophisticated models. Third, we know the "true" data, (i.e. the real data). By comparing the real data with synthetic data generated under different models, we may be able to identify certain assumptions of our model that are especially important.

Figure 2 shows the mean root age and 95% confidence interval for a series of synthetic data analyses carried out using TraitLab. Results to the left of the vertical line are for the finite-sites model of method 1, whilst those to the right are for method 2, fitting the stochastic-Dollo model. These results and the evolutionary models used to generate each set of data are summarized in Table 4.

Both the methods were able to reconstruct the age of data synthesized under their respective models (SF1, SD1). More interestingly, TraitLab allows data to be simulated under a number of other evolutionary scenarios. Data was synthesized on a number of different trees; however, the results presented here all use the same tree, chosen from the posterior distribution in RD16. The true age was 8680 years.

First, we investigated the effect of borrowing on divergence time estimates. We generated a series of synthetic data sets using models of evolution that allow for horizontal as well as vertical transmission of cognates between random pairs of languages. As with the standard models, cognates evolve through time along each lineage according to the stochastic-Dollo model of word birth/death. In addition, however, for any given time interval, cognates can be borrowed from one lineage to another randomly selected existing lineage with a certain probability. By varying this probability we can simulate the effect of different rates of borrowing. Even relatively high rates of borrowing, at 20% of the cognate death rate (SF2 and SD3), had only a minimal effect on divergence time estimates, causing a slight underestimation of ages in both methods of analysis. As the rate of borrowing increased the extent of underestimation also increased. For a borrowing rate of 100%, date estimates were reduced by 30% (SF3 and SD4).

Second, it may be unrealistic to assume that words can be borrowed from any language to any other language with equal probability. For example, Western-European languages may be far more likely to borrow from other Western-European languages

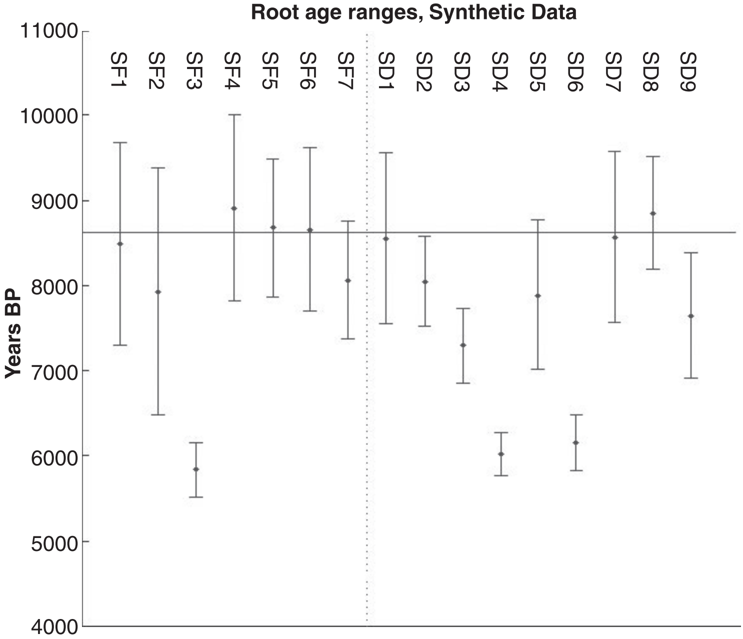


Figure 2. Mean root age and 95% confidence interval for a series of synthetic data analyses carried out using TraitLab. Data was generated under a number of models of evolution on a tree with a root age of 8680 years BP (indicated by the horizontal line) chosen from the posterior distribution in RD16. Results to the left of the vertical line are for method 1, fitting the finite sites model, whilst those to the right are for method 2, fitting the stochastic-Dollo model using TraitLab.

than from Iranian languages. For this reason, we also synthesized data with borrowing limited to local areas of the tree. This was achieved by restricting the borrowing process to only those languages that had diverged within a certain threshold cut-off time. If, for example, a 4000 year threshold was used, at any given time interval on the tree, lineages that had been separated for over 4000 years could not borrow words between them. This has the

Table 4. Summary of results shown in figure 2 for synthetic data analyses, including the mean and standard deviation for the estimated age at the root of the tree on which data was synthesized. Data was synthesized on a tree chosen from the posterior distribution in RD16. The true age was 8680 units.

Analysis	μ	S.D.	Synthetic data model
SF1	8488	595	Finite-sites
SF2	7925	725	Stochastic-Dollo with 20% global borrowing
SF3	5838	160	Stochastic-Dollo with 100% global borrowing
SF4	8911	545	Stochastic-Dollo with 20% local borrowing, 4000 years
SF5	8677	404	Dependent model
SF6	8660	479	Dependent model with 20% local borrowing, 1000 years
SF7	8061	350	Dependent model with 20% local borrowing, 1500 years
SD1	8558	500	Stochastic-Dollo model
SD2	8054	262	Stochastic-Dollo Model with 10% global borrowing
SD3	7291	220	Stochastic-Dollo Model with 20% global borrowing
SD4	6021	123	Stochastic-Dollo Model with 100% global borrowing
SD5	7891	443	Stochastic-Dollo Model with 100% local borrowing, 500 years
SD6	6157	161	Stochastic-Dollo Model with 20% local borrowing 4000 yrs
SD7	8567	499	Dependent model
SD8	8853	331	Dependent model with 20% local borrowing, 1000 years
SD9	7642	370	Dependent model with 20% local borrowing, 1500 years

effect of eliminating borrowing between the most distantly related languages, such as between the Germanic and Iranian languages. Using a 4,000 year threshold and 20% borrowing rate caused method 2 (SD6), but not method 1 (SF4), to underestimate divergence times. More localized borrowing tended to result in more accurate date estimates. Interestingly, using a much lower borrowing threshold, such as 500 years, and relatively high rates of borrowing allows us to approximate the effects of dialect chain divergence, where languages remain in contact for a period after beginning to separate. Again, this had little effect on divergence time estimates (SD5).

Finally, data were generated under a dependent model of cognate evolution to test the effect of violations of the independence assumption. The dependent model of evolution models multiple subsets of cognates representing meaning categories. Under this model, cognates can evolve independently between meaning categories but are subject to constraints within each meaning category. Each language must always have at least one cognate in each meaning category. In addition, by varying parameters controlling the expected vocabulary size relative to the number of meaning categories, we can alter the expected number of words within each meaning category in each language. The resulting model is dependent in that each language must have at least one word for each meaning category. However, the converse feature of the model proposed by Evans *et al.* (in press) is not explicitly modeled: if a word is present, other words of the same meaning are born and die independently. It happens that the probability distribution for the number of words in each category (Poisson, conditioned away from zero) does decline from a maximum at one word (for the parameters we chose when we generate synthetic data). This dependent model may more accurately reflect the true process of language evolution. Both methods perform very well at recovering the true age (SF5, SD7). This remains the case even if we introduce borrowing (SF6, SF7, SD8, SD9).

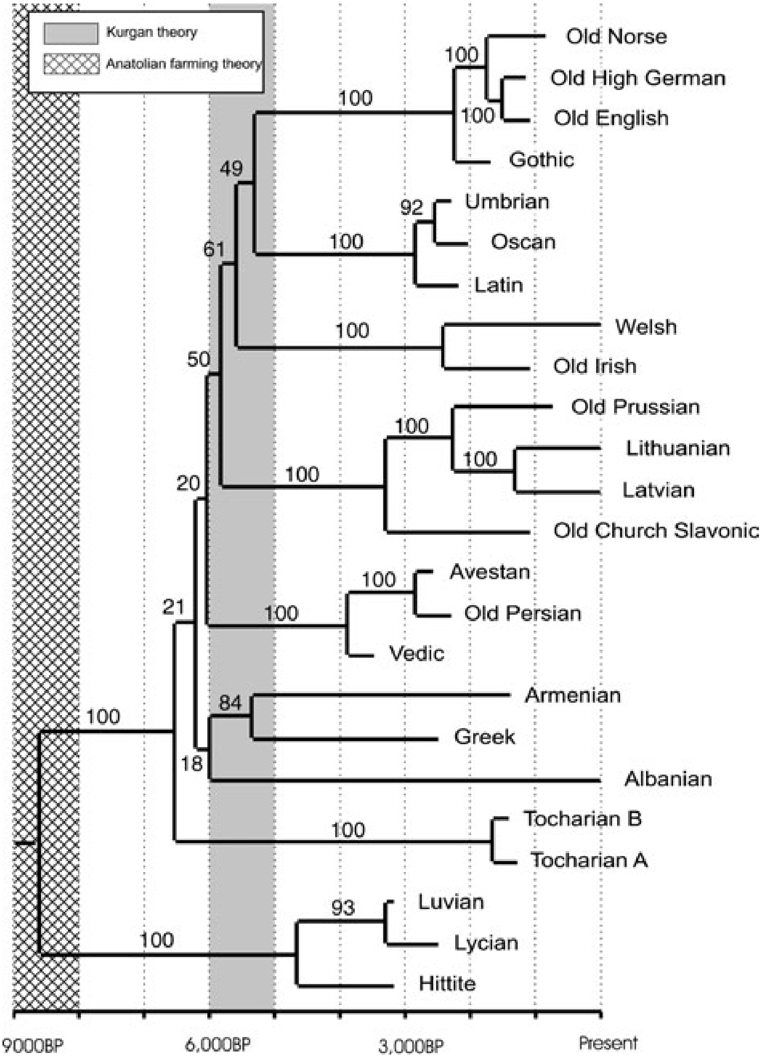
9. Discussion

The divergence time estimates derived from the two independent lexical data sets using two very different models of word evolution are strikingly consistent. The findings of Gray & Atkinson (2003) seem to be robust to the choice of languages sampled, to the meaning categories analysed, to who makes vocabulary assignments and cognacy judgements, and even to the age of the languages sampled. The fact that ancient and modern data sets have produced similar date estimates strongly supports the notion that the process of language evolution itself is sufficiently constrained and robust to human socio-cultural change as to make date estimates based on lexical comparison a feasible possibility. This is also evidence against the suggestion that contemporary borrowing could bias the age estimates.

Of course, one problem could be with the methodology. It is therefore impressive that two different methods using very different models give the same result. We were able to address the principal concerns that have been raised about model misspecification using synthetic data. The models were found to be robust to key criticisms of borrowing and independence. Significantly, if the age implied by the Kurgan hypothesis were the true age of Indo-European, model misspecification would have to cause us to over-estimate the age of the common ancestor by a factor of one and one half in order for us to find support for the 8,000 BP to 9,500 BP age range implied by the Anatolian theory. None of the types of model misspecification tested here produced appreciable overestimation. Analyses of data synthesized under models incorporating various degrees of global and local borrowing produced progressively greater underestimation of age estimates as the degree and extent of the borrowing was increased. Interestingly, whilst the finite-sites model performed relatively well in reconstructing the root age from data synthesized under a stochastic-Dollo model with borrowing (e.g. SF2 & SF4), it could not reconstruct reasonable branch-lengths from data synthesized under the strict Dollo model, with no borrowing. The time reversible nature of the model appears to allow it to effectively accommodate for homoplasy due to borrowing, however, when applied to homoplasy-free synthetic data the model has difficulty. When data was generated under a "dependent" model of evolution both models were able to estimate the root age relatively accurately. This suggests that whilst these models assume that the evolution of characters is independent, they are in fact robust to violations of this assumption. Certainly, there is no evidence to suggest that

---

Figure 3. Majority-rule consensus tree from the initial Bayesian MCMC sample of 1,000 trees based on the Ringe *et al.* (2002) data. Values above each branch indicate uncertainty (posterior probability) in the tree as a percentage. Branch-lengths are proportional to time. Shaded bars represent the age range proposed by the two main theories – the Anatolian theory (grey bar) and the Kurgan theory (hatched bar). The basal age (8,680 BP) supports the Anatolian theory. While consensus trees are a useful visual aid, it is known that the branch support values become unreliable when there is substantial model misspecification. The consensus network provided in the supplementary material is a better representation of the results, although perhaps more difficult to interpret.


Kurgan theory
Anatolian farming theory
Old Norse
Old High German
Old English
Gothic
Umbrian
Oscan
Latin
Welsh
Old Irish
Old Prussian
Lithuanian
Latvian
Old Church Slavonic
Avestan
Old Persian
Vedic
Armenian
Greek
Albanian
Tocharian B
Tocharian A
Luvian
Lycian
Hittite
100
49
61
50
20
21
18
84
92
93
9000BP
6,000BP
3,000BP
Present

violations of the independence assumption might cause us to grossly over-estimate ages.

A key aspect of our approach is that we can make date estimates without having to postulate just one phylogeny. Instead, inferences about divergence times are made on the basis of the Bayesian sample distribution of trees, allowing us to quantify the phylogenetic uncertainty implicit in our date estimates. We can, however, use a consensus tree or consensus network (Holland & Moulton, 2003) as a visualization tool. A consensus network is provided in the supplementary material. Figure 3 shows a consensus tree constructed using the distribution of trees from the standard RF1 analysis of the Ringe *et al.* (2002) data. Branch-lengths are proportional to time. As well as tree topology and branch length information, the consensus tree shows the degree of support for each sub-clade within the phylogeny, expressed in the form of the "posterior probability" of a clade — the percentage of time that the clade appears in the Bayesian MCMC sample distribution. A value of 100, for example, indicates that the clade occurs in all sampled trees. Lower values indicate an increasing degree of statistical uncertainty. It is evident from figure 3 that the two theories of Indo-European origin may not, in fact, be mutually exclusive – a possibility identified by Cavalli-Sforza *et al.* (1994). Whilst the basal age (8,680 BP) supports the Anatolian theory of Indo-European origin, there is a period of rapid divergence during the hypothesized time of the Kurgan expansion, between 5,000 BP and 6,000 BP. Gray & Atkinson (2003) note a similar pattern in their analysis of the Dyen *et al.* (1997) data.

10. CONCLUSION

The well-known criticisms of glottochronology have led many researchers to reject the possibility of estimating dates from lexical data. Any approach that attempts to turn words into dates is dismissed as attempting the impossible – trying to turn water into wine. Here we have shown that estimating divergence time confidence intervals from lexical data is far from impossible or miraculous. New statistical tools from evolutionary biology enable us to estimate phylogeny and divergence times without falling victim to the pitfalls of glottochronology. The availability of these methods means that it is no longer valid to dismiss all attempts at estimating divergence dates simply because Swadesh's approach was flawed. If used sensibly these new methods offer a powerful set of tools with the potential to resolve some of the long-standing debates in historical linguistics.

Quentin Atkinson and Russell Gray
Department of Psychology,
University of Auckland,
Private Bag 92019,
Auckland 1020,
New Zealand.
Email: rd.gray@auckland.ac.nz

Geoff Nicholls and David Welch
Department of Mathematics,
University of Auckland,
Private Bag 92019,
Auckland 1020,
New Zealand.

REFERENCES

ADAMS, D. Q., 1999. *A Dictionary of Tocharian B* (Leiden Studies in Indo-European 10). Amsterdam: Rodopi. Available via online database at *S. Starostin and A. Lubotsky* (Eds.) Database Query to A dictionary of Tocharian B. http://iiasnt.leidenuniv.nl/ied/index2.html.

ATKINSON, Q. D. & GRAY, R. D., in press[a]. 'Are accurate dates an intractable problem for historical linguistics?', in In C. Lipo, M. O'Brien, S. Shennan & M. Collard (eds.), *Mapping our Ancestry: Phylogenetic Methods in Anthropology and Prehistory*, Chicago: Aldine.

ATKINSON, Q. D. & GRAY, R. D., in press[b]). 'How old is the Indo-European language family? Illumination or more moths to the flame?', in Peter Forster, Colin Renfrew and James Clackson (eds.), *Phylogenetic methods and the prehistory of languages*, Cambridge: McDonald Institute for Archaeological Research.

BERGSLAND, K., & VOGT, H., 1962. 'On the validity of glottochronology', Current Anthropology 3, 115–153.

BLUST, R., 2000. 'Why lexicostatistics doesn't work: the 'universal constant' hypothesis and the Austronesian languages', in Colin Renfrew, April McMahon and Larry Trask (eds.), *Time Depth in Historical Linguistics*, Cambridge: The McDonald Institute for Archaeological Research, 311–332.

Bryant, D., Filimon, F. & Gray, R.D., in press. 'Untangling our past: Pacific settlement, phylogenetic trees and Austronesian languages', in R. Mace, C. Holden and S. Shennan (eds.), *The Evolution of Cultural Diversity: Phylogenetic approaches*, London: UCL Press.

Bryant, D., & Moulton, V., 2002. 'NeighborNet: An agglomerative method for the construction of planar phylogenetic networks', Workshop in Algorithms for Bioinformatics, Proceedings 2002: 375–391.

Campbell, Lyle, 2004. *Historical Linguistics: An introduction*. 2nd edition. Edinburgh: Edinburgh University Press.

Cavalli-Sforza, Luigi Luca, Menozzi, Paolo & Piazza, Alberto, 1994. *The History and Geography of Human Genes*, Princeton, N.J.: Princeton University Press.

Dyen, Isidore, Kruskal, Joseph B. & Black, Paul, 1992. *An Indoeuropean Classification: A Lexicostatistical Experiment*. Philadelphia: American Philosophical Society, Transactions 82(5).

Dyen, Isidore, Kruskal, Joseph B. & Black, Paul, 1997. FILE IE-DATA1. Available at http://www.ntu.edu.au/education/langs/ielex/IE-DATA1.

Evans, S. N., Ringe, Donald & Warnow, Tandy, in press. 'Inference of divergence times as a statistical inverse problem', in James Clackson, Peter Forster and Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*. Cambridge: McDonald Institute for Archaeological Research.

Farris, J. S., 1977. 'Phylogentic analysis under Dollo's Law', Systematic Zoology 26, 77–88.

Gamkrelidze, T. V., & Ivanov, V.V., 1995. *Indo-European and the Indo-Europeans: A Reconstruction and Historical Analysis of a Proto-Language and Proto-Culture*, Berlin: Mouton de Gruyter.

Garrett, Andrew, in press. 'Convergence in the formation of Indo-European subgroups: Phylogeny and chronology', in James Clackson, Peter Forster and Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, Cambridge: McDonald Institute for Archaeological Research.

Geyer, C.J., 1992. 'Practical Markov chain Monte Carlo', Statistical Science 7, 473–511

Gimbutas, M., 1973a. 'Old Europe c. 7000–3500 B.C., the earliest European cultures before the infiltration of the Indo-European peoples', Journal of Indo-European Studies 1, 1–20.

Gimbutas, M. 1973b. 'The beginning of the Bronze Age in Europe and the Indo-Europeans 3500–2500 B.C.', Journal of Indo-European Studies 1, 163–214.

Gray, R. D. & Atkinson, Q.D., 2003. 'Language-tree divergence times support the Anatolian theory of Indo-European origin', Nature 426, 435–9.

Guterbock, H. G. & Hoffner, H.A., 1986. *The Hittite dictionary of the Oriental Institute of the University of Chicago*. Chicago: The Institute.

Hoffner, H. A., 1967. *An English-Hittite Dictionary*. New Haven: American Oriental Society.

Holland, B., & Moulton, V., 2003. 'Consensus networks: a method for visualising incompatibilities in collections of trees', in G. Benson and R. Page (eds.), *Algorithms in bioinformatics, WABI 2003*, Berlin: Springer-Verlag, 165–176.

Huelsenbeck, J. P., & Ronquist, F., 2001. 'MRBAYES: Bayesian Inference of Phylogeny', Bioinformatics 17, 754–755.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J.P., 2001. 'Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology', Science 294, 2310–2314.

Huson, D. H., 1998. 'SplitsTree: Analyzing and visualizing evolutionary data', Bioinformatics 1.4, 68–73.

Jefferson, T., 1781–1782. reprint of *Notes on the State of Virginia. M. D. Peterson*, ed.). 1984. New York: Library of America, Literary classics of the United States.

Kuhner, M. K., & Felsenstein, J., 1994. 'A Simulation Comparison of Phylogeny Algorithms under Equal and Unequal Evolutionary Rates', Molecular Biology and Evolution 11, 459–468.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E., 1953. 'Equations of State Calculations by Fast Computing Machines', Journal of Chemical Physics 21, 1087–1091.

Nicholls, G., & Gray, R., in press. 'Quantifying uncertainty in a stochastic dollo model of vocabulary evolution', in James Clackson, Peter Forster and Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, Cambridge: McDonald Institute for Archaeological Research.

Pagel, M., 1997. 'Inferring Evolutionary Processes from Phylogenies', Zoologica Scripta 26, 331–348.

Renfrew, C., 1987. *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Cape.

Rexová, K., Frynta, D., & Zrzavy, J., 2003. 'Cladistic analysis of languages: Indo-European classification based on lexicostatistical data', Cladistics 19, 120–127.

Ringe, Don, Warnow, Tandy & Taylor, Ann, 2002. 'Indo-European and Computational Cladistics', Transactions of the Philological Society 100, 59–129.

Sanderson, M., 2002a. 'Estimating absolute rates of evolution and divergence times: A penalized likelihood approach', Molecular Biology and Evolution 19, 101–109.

Sanderson, M., 2002b. *R8s, Analysis of Rates of Evolution, version 1.50*. http://ginger.ucdavis.edu/r8s/.

Steel, M., Hendy, M. & Penny, D., 1988. 'Loss of information in genetic distances', Nature 333, 494–495.

Swadesh, M., 1952. 'Lexico-statistic dating of prehistoric ethnic contacts', American Philosophical Society, Proceedings 96, 453–463.

Swadesh, M., 1955. 'Towards greater accuracy in lexicostatistic dating', International Journal of American Linguistics 21, 121–137.

Swofford, D. L., Olsen, G. J., Waddell, D. J. & Hillis, D. M., 1996. 'Phylogenetic Inference', in D. M. Hillis, C. Moritz, and B. K. Marble (eds.), *Molecular Systematics*, 2nd edition. Sunderland, Mass.: Sinauer, 407–514.

Tischler, J., 1973. *Glottochronologie und Lexicostatistik*. Innsbruck: Innsbrucker Verlag.

Tischler, J., 1997. *Hethitisch-Deutsches Worterverzeichnis*. Dresden: Probedruck.