

The nuclear family
in the history of Indo-European-speaking societies:
a phylogenetic comparative approach
(supplementary materials)

Laura Fortunato

Institute of Cognitive and Evolutionary Anthropology

University of Oxford

64 Banbury Road, Oxford OX2 6PN, UK

`laura.fortunato@anthro.ox.ac.uk`

+44 (0)1865 284971

Santa Fe Institute

1399 Hyde Park Road

Santa Fe, NM 87501, USA

DRAFT

PLEASE DO NOT CITE OR CIRCULATE

S1 Data and methods

S1.1 Phylogenetic comparative analysis

I used data on marriage and residence strategies from the *Ethnographic Atlas* (*EA*) database (Gray 1999; Murdock 1967) for a sample of 27 societies speaking Indo-European (IE) languages (Section S1.1.1). The cross-cultural data were mapped onto Pagel et al.'s (2007) sample of phylogenetic trees (Section S1.1.2). On the cross-cultural data and tree sample, I used the phylogenetic comparative method developed by Pagel and colleagues (Pagel and Meade 2005, 2006; Pagel et al. 2004) to test the hypothesis of co-evolution between monogamous marriage and neolocal residence (Section S1.1.3).

S1.1.1 Cross-cultural data

I obtained the cross-cultural sample by matching societies in the *EA* database (Gray 1999; Murdock 1967) with speech varieties in the Dyen et al.'s (1992) database, as described in Fortunato (2011c). Variable identifiers in this section follow Gray's (1999) *EA* codebook.

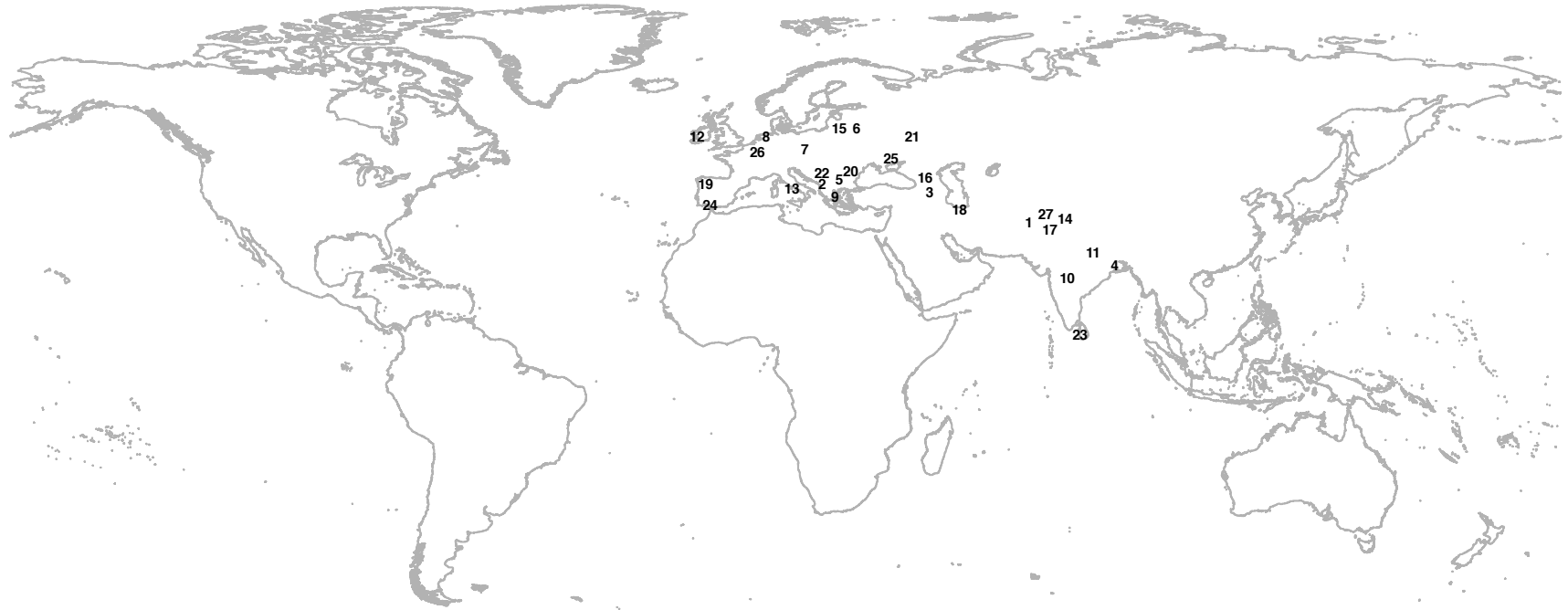
The method used for phylogenetic comparative analysis requires data in binary form (Section S1.1.3). For marriage strategy I coded societies as polygynous (state 0) or monogamous (state 1) based on *EA* variable 9, as described in Fortunato (2011a). For residence strategy I first coded societies as neolocal, uxori-local, or viri-local, separately for prevailing and alternative modes of residence, based on *EA* variables 12 and 14, as described in Fortunato (2011b). Next, I combined this information into one binary variable, with societies coded as non-neolocal (state 0) or neolocal (state 1). Theoretically, this coding is justified on the grounds that any degree of neolocality is assumed to weaken extended family organization, even where the prevailing mode of residence is non-neolocal (Murdock 1949, p. 208).

This produced a sample with monogamy present in 15 (93.8%) of 16 neolocal societies and absent in 1 (6.2%), and with monogamy present in 3 (27.3%) of 11 non-neolocal societies and absent in 8 (72.7%) (Table S1 and Figures S1b and S1c).

Table S1: Recoded *EA* data

Key ¹	Speech variety ²	<i>EA</i> identifiers ³		Marriage strategy ⁴	Residence strategy ⁵
		Code	Name		
1	Afghan	Ea11	Afghans	Polygynous	Non-neolocal
2	Albanian G	Ce1	Gheg	Polygynous	Non-neolocal
3	Armenian Mod	Ci10	Armenians	Monogamous	Non-neolocal
4	Bengali	Ef2	Bengali	Polygynous	Non-neolocal
5	Bulgarian	Ch5	Bulgarians	Monogamous	Neolocal
6	Byelorussian	Ch6	Byelorussians	Monogamous	Neolocal
7	Czech	Ch3	Czechs	Monogamous	Neolocal
8	Dutch List	Cg1	Dutch	Monogamous	Neolocal
9	Greek MD	Ce7	Greeks	Monogamous	Neolocal
10	Gujarati	Ef9	Gujarati	Polygynous	Non-neolocal
11	Hindi	Ef11	Uttar Pradesh	Polygynous	Non-neolocal
12	Irish B	Cg3	Irish	Monogamous	Neolocal
13	Italian	Ce5	Neapolitans	Monogamous	Neolocal
14	Kashmiri	Ef8	Kashmiri	Polygynous	Non-neolocal
15	Lithuanian ST	Ch9	Lithuanians	Monogamous	Neolocal
16	Ossetic	Ci6	Osset	Polygynous	Non-neolocal
17	Panjabi ST	Ea13	Punjabi	Monogamous	Non-neolocal
18	Persian List	Ea9	Iranians	Polygynous	Neolocal
19	Portuguese ST	Ce2	Portuguese	Monogamous	Neolocal
20	Rumanian List	Ch10	Romanians	Monogamous	Neolocal
21	Russian	Ch11	Russians	Monogamous	Neolocal
22	Serbocroatian	Ch1	Serbs	Monogamous	Neolocal
23	Singhalese	Eh6	Sinhalese	Monogamous	Non-neolocal
24	Spanish	Ce6	Spaniards	Monogamous	Neolocal
25	Ukrainian	Ch7	Ukrainians	Monogamous	Neolocal
26	Walloon	Cg5	Walloons	Monogamous	Neolocal
27	Waziri	Ea2	Pathan	Polygynous	Non-neolocal

¹ Refers to the numbers in Figure S1a.² After Dyen et al. (1992); see text for details.³ After Gray (1999); see text for details.⁴ Recoded from *EA* variable 9; see text for details.⁵ Recoded from *EA* variable 12 for prevailing mode and 14 for alternative mode; see text for details.



(a) Key

Figure S1: Map illustrating the geographic distribution of the recoded *EA* data in Table S1. Numbers correspond to entries in the relevant column in Table S1.



(b) Marriage strategy

Figure S1: Map illustrating the geographic distribution of the recoded *EA* data in Table S1. Colours express the marriage strategy (white: monogamous; black: polygynous), as per Figure 3a in the main text.



(c) Residence strategy (binary)

Figure S1: Map illustrating the geographic distribution of the recoded *EA* data in Table S1. Colours express the residence strategy (white: neolocal; black: non-neolocal), as per Figure 3b in the main text.

S1.1.2 Tree sample

I used Pagel et al.’s (2007) posterior probability sample of 750 phylogenetic trees to represent how societies in the cross-cultural sample are related by way of descent from a common ancestor. Trees are present in the sample in proportion to their posterior probabilities. The posterior probability of a tree can be interpreted as the probability that the tree is correct (Huelsenbeck et al. 2001). Pagel et al. (2007) inferred the posterior probability distribution of trees from the linguistic database using the Bayesian Markov chain Monte Carlo (MCMC) phylogenetic tree-building method developed by Pagel and Meade (2004). The tree-building analysis used linguistic data for 87 speech varieties (Pagel et al. 2007).

I obtained the tree sample from the authors. I “pruned” the trees to retain only the speech varieties corresponding to the 27 societies in the cross-cultural sample (Section S1.1.1), plus the outgroup Hittite. Hittite was assigned no data on marriage and residence for the purpose of the comparative analysis (Section S1.1.3). A comprehensive description of related data, methods, and procedures is in Fortunato (2011a,c).

S1.1.3 Comparative analysis

Co-evolutionary analysis using the phylogenetic comparative method developed by Pagel and colleagues (Pagel and Meade 2005, 2006; Pagel et al. 2004) is performed using `BayesDiscrete`, available as part of the `BayesTraits` package (Pagel and Meade n.d.) from <http://www.evolution.rdg.ac.uk/BayesTraits.html>. Unless otherwise specified, the information in this section is based on Pagel and Meade (2005, 2006), Pagel et al. (2004), and on the `BayesTraits` manual (Pagel and Meade n.d.). A detailed non-technical discussion of the method and its application is in Fortunato (2009, Appendix C).

Models of trait evolution `BayesDiscrete` uses the cross-cultural data and tree sample to estimate parameters in the model of trait evolution for two binary traits. In this case the comparative data consists of one trait representing the marriage strategy, with states 0 (polygynous) and 1 (monogamous), and one trait representing the residence strategy, with states 0 (non-neolocal) and 1 (neolocal) (Section S1.1.1). This produces four combinations of states for the two traits, e.g. combination 2 (0, 1) for polygynous marriage with neolocal residence, as shown in Figure 1 in the main text. For ease of reference, the figure is included in this section

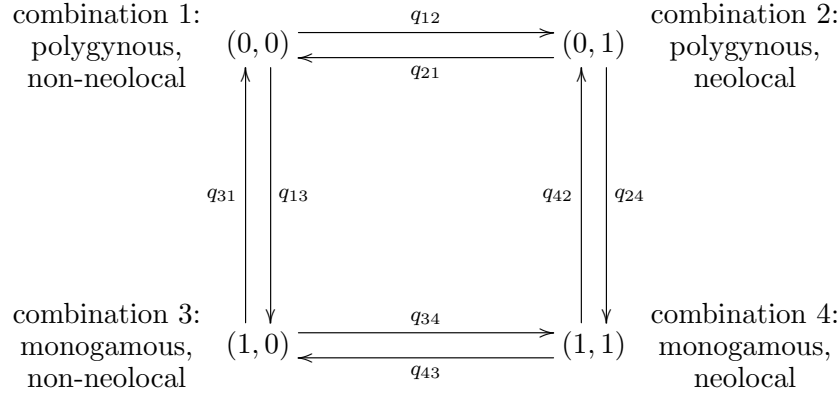


Figure S2: Rate parameters describing the evolution of marriage and residence strategies.

as Figure S2.

A series of transition rate parameters q_{ij} describes the joint evolution of the two traits on a tree. These rate parameters measure the instantaneous rate of change from combination “i” to combination “j”, which corresponds to the rate of change between two states of one trait while holding the state of the other trait constant. For example, rate parameter q_{12} describes transitions from combination 1 (0, 0) to combination 2 (0, 1), that is, changes from non-neolocal to neolocal residence against a background of polygynous marriage (Figure S2). Similarly, rate parameter q_{34} describes transitions from combination 3 (1, 0) to combination 4 (1, 1), that is, changes from non-neolocal to neolocal residence against a background of monogamous marriage. Rate parameters describing simultaneous transitions in the two traits, corresponding to the diagonals of the diagram in Figure S2, are set to zero: this amounts to assuming that the probability of the two traits changing simultaneously is negligibly small.

The rate parameters are used to define the probabilities of the corresponding changes, the probabilities of the two states at internal nodes on the tree, and the likelihood of the data. Tips that are not assigned comparative data for a given trait are treated in the likelihood calculations as taking either state for the trait with equal probability. This is the case for the outgroup Hittite (Section S1.1.2).

Independent vs. dependent trait evolution The eight rate parameters yield 21,146 model categories, in which two or more rate parameters are set to take equal or distinct positive values, or one or more of them can be set to zero. Of the 21,146 possible categories, 51 (0.24%) conform

to independent trait evolution. In these categories the rate of change in one trait is not affected by the state of the other trait, that is, $q_{12} = q_{34}$, $q_{13} = q_{24}$, $q_{21} = q_{43}$, and $q_{31} = q_{42}$ (Figure S2). All other model categories conform to dependent trait evolution, because they assign rates within one or more of these pairs to different rate classes; this means that the rate of change in one trait does depend on the state of the other trait. For example, $q_{12} = q_{34}$ specifies that the rate of change from non-neolocal to neolocal residence does not depend on whether marriage is polygynous or monogamous; $q_{12} \neq q_{34}$ indicates that it does.

Evidence for dependent vs. independent evolution I used BayesDiscrete in Bayesian reversible jump (RJ) MCMC mode. In this mode BayesDiscrete uses RJ-MCMC methods to estimate the posterior probability distributions of rate parameters, of ancestral states at internal nodes on a tree, and of the possible models of trait evolution specified by the rate parameters. The posterior probability of a parameter value is a quantity proportional to its likelihood of having produced the observed data, and represents the probability of the parameter value given the data and model of trait evolution. Schematically, the posterior probability distributions are estimated by running RJ-MCMC chains that sample states in the model of trait evolution in proportion to their posterior probability, across trees in the tree sample. A state in the model consists of model category, values of the rate parameters, and ancestral state probabilities. Combining estimates over the sample produced by a chain amounts to “averaging” inferences over uncertainty in the phylogeny, in the parameters of the model of trait evolution, and in the model itself.

The proportions of model categories conforming to dependent vs. independent evolution sampled by a chain provide a direct estimate of the posterior probability that the traits evolved dependently vs. independently. The theoretical prior odds is $(21,146 - 51)/51 = 413.63$: this means that if all model categories had equal posterior probability, the chain would sample categories conforming to dependent evolution approximately 414 times more frequently than categories conforming to independent evolution. The Bayes factor, obtained as the ratio of the posterior odds returned by the chain to the prior odds, provides an estimate of the posterior probability of the two evolutionary models. The Bayes factor for dependent vs. independent evolution is denoted B_{DI} . Values of $B_{DI} > 1$ represent evidence for dependent trait evolution, values of $B_{DI} < 1$ evidence for independent trait evolution. Equivalently, the Bayes factor can

Table S2: Scales for interpretation of the Bayes factor*

$ B_{DI} $	$ 2\log_e(B_{DI}) $	Evidence for evolutionary model
1 to 3	0 to 2	Weak
3 to 12	2 to 5	Positive
12 to 150	5 to 10	Strong
> 150	> 10	Very strong

* After Raftery (1996, p. 165).

be expressed on a logarithmic scale, with $2\log_e(B_{DI}) > 0$ representing evidence for dependent trait evolution and $2\log_e(B_{DI}) < 0$ evidence for independent trait evolution. The strength of the evidence in favour of one model over the other is assessed against the scales in Table S2.

This implementation of the Bayes factor assumes all model categories to be equally likely, and it may thus result in conservative estimates of support for dependent trait evolution. This is because, in practice, the processes underlying the evolution of the two traits may involve fewer than the eight transitions specified by rate parameters q_{ij} . If the “true” number of rate parameters is less than eight, then the number of possible model categories will be less than 21,146. Since model categories conforming to independent evolution require a maximum of four distinct rate classes, they will represent a higher proportion of all the possible categories, compared to the theoretical value (51 of 21,146). As a result, the “true” prior odds will be lower than the theoretical value of $(21,146 - 51)/51 = 413.63$. Consequently, the posterior to prior odds Bayes factor obtained using this value will be lower than the “true” Bayes factor, potentially underestimating support for dependent evolution.

This issue can be circumvented through an alternative implementation of the Bayes factor. This involves running a separate chain that is constrained to visit model categories conforming to independent evolution. The Bayes factor is obtained by comparing the posterior probability distribution of $\log_e(\text{likelihood})$ values returned by this chain to the distribution returned by an unconstrained chain (having removed visits to model categories conforming to independent evolution). In this case, $2\log_e(B_{DI})$ is approximated as twice the difference between $\log_e[H(\text{likelihood})]$ for the unconstrained chain and $\log_e[H(\text{likelihood})]$ for the constrained chain, where $\log_e[H(\text{likelihood})]$ is the natural logarithm of the harmonic mean of the likelihood values. The strength of the evidence in favour of one model over the other is assessed against the logarithmic scale in Table S2. Because $\log_e[H(\text{likelihood})]$ may be unstable, Pagel

and Meade (n.d.) recommend taking a conservative approach, which disregards any evidence for either model given by $|2\log_e(B_{DI})| < 2$.

Implementation I estimated support for dependent trait evolution using both implementations of the Bayes factor. This involved running two sets of analyses: one, termed RJ-unconstrained, in which chains were free to visit all possible model categories, and one, termed RJ-constrained, in which chains were constrained to visit categories conforming to independent trait evolution. Each set of analyses comprised five separate chains started from random seeds.

I determined the RJ-MCMC chain specifications through preliminary maximum-likelihood and MCMC runs. These specifications ensure that the RJ-MCMC chains sample parameter space adequately and ultimately converge to the posterior probability distribution of states in the model of trait evolution. I ran the RJ-MCMC chains for 10^8 iterations, sampling every 10^3 , with an additional burn-in of 10^6 , and rate deviation set to 60. Under maximum-likelihood the covarion model for trait evolution improved the mean $\log_e(\text{likelihood})$ by between 0.6 and 3.7 units, depending on the model specifications, and was therefore implemented by all RJ-MCMC chains; this model allows rates to vary within and between branches of a tree. All chains used a uniform prior on the models and an exponential prior on the rate parameters; the mean of the exponential prior was seeded from a uniform hyperprior on the interval 0–10. The shape of the prior distribution (exponential or gamma) and the interval of the hyperprior (0–10 or 0–20) only marginally affected the mean of the posterior probability distributions of $\log_e(\text{likelihood})$ values and of ancestral states, returning qualitatively similar results.

For each set of analyses, convergence to the posterior probability distribution of states in the model of trait evolution was assessed by comparing the samples returned by the separate chains, through visual inspection of (i) time-series plots of $\log_e(\text{likelihood})$ values, (ii) the posterior probability distributions of model categories, and (iii) the average deviation of parameter estimates across runs. The near-independence of sampling events was judged from the autocorrelation of the $\log_e(\text{likelihood})$ values of successive states sampled by the chains. In all cases, these diagnostics indicated that the chains sampled the target distributions adequately.

S2 Results

S2.1 Phylogenetic comparative analysis

For each sets of analyses, I compared $\log_e[H(\text{likelihood})]$ values for the states in the model of trait evolution sampled by the separate chains at convergence; I present results for the chain that returned the median value of $\log_e[H(\text{likelihood})]$.

I discuss separately the posterior probability distribution of model categories returned by the RJ-unconstrained chain (Section S2.1.1) and the estimates of support for dependent trait evolution obtained from the two implementations of the Bayes factor (Section S2.1.2).

S2.1.1 Posterior probability distribution of model categories

At convergence, the RJ-unconstrained chain sampled states in the model of trait evolution with mean $\log_e(\text{likelihood}) \pm \text{SD} = -18.86 \pm 1.34$ (range: -30.67 to -14.65 ; auto-correlation coefficient: $r = 0.055$; mean acceptance rate: 33.2%).

The chain sampled 2,855 model categories, with rates assigned to 1.86 ± 0.45 non-zero classes (mean $\pm$ SD; range: 1–4 non-zero classes). This indicates that the processes underlying the evolution of the two traits likely involved fewer than the eight transitions specified by rate parameters q_{ij} . For example, Table S3 shows the ten model categories sampled most frequently by the chain; the hundredth most frequently sampled category is included for comparison. The top ten categories, which account for 18.2% of the 10^5 sampled points, include only one or two non-zero rate classes. Of these, the fifth sets all rates equal to each other, conforming to independent evolution, while the other nine conform to dependent evolution.

S2.1.2 Estimation of support for dependent trait evolution

Of the 10^5 points sampled by the RJ-unconstrained chain, 98.67% corresponded to model categories conforming to dependent evolution and 1.33% to categories conforming to independent evolution, yielding a posterior odds of 74.36. Comparison of the posterior odds to the prior odds yields $B_{\text{DI}} = \frac{74.36}{413.63} = 0.18$ or, on a logarithmic scale, $2 \log_e(B_{\text{DI}}) = -3.43$. This corresponds to no evidence for dependent trait evolution or, equivalently, to positive evidence for independent trait evolution (Table S2).

Table S3: Rate classes and frequencies for selected model categories

Model category	Rate class [*]								Frequency [†]	
	q_{12}	q_{13}	q_{21}	q_{24}	q_{31}	q_{34}	q_{42}	q_{43}	Absolute	Relative cumulative
1	0	0	0	Z	0	0	Z	0	5471	0.055
2	0	0	0	0	0	0	Z	0	4636	0.101
3	Z	0	0	0	0	0	0	Z	1987	0.121
4	0	1	1	Z	1	1	Z	0	1069	0.132
5	0	0	0	0	0	0	0	0	912	0.141
6	0	1	1	1	1	1	Z	0	864	0.149
7	Z	0	0	0	0	0	0	0	837	0.158
8	0	0	0	0	0	Z	0	0	832	0.166
9	0	0	0	Z	0	0	0	0	790	0.174
10	0	0	0	Z	0	0	Z	1	779	0.182
⋮										⋮
100	0	1	0	1	0	1	Z	0	242	0.531

^{*} “Z” denotes rates assigned to the zero class. “0” and “1” denote two distinct non-zero rate classes; rates with the same value are assigned to the same non-zero rate class. For example, under model category 1 q_{12} , q_{13} , q_{21} , q_{31} , q_{34} , and q_{43} are assigned to the same non-zero rate class, that is, they take the same positive value, while q_{24} and q_{42} are assigned to the zero rate class, that is, they are set to zero.

[†] The relative cumulative frequency of a model category is obtained by summing the absolute frequency of sampled points in the model category to the absolute frequencies of sampled points in all preceding categories, and then dividing by the total number of sampled points in all categories [e.g. $(5471 + 4636)/10^5 = 0.101$ for model category 2].

Because the RJ-unconstrained chain sampled a restricted sub-set of all possible model categories, with no categories including five or more distinct non-zero rate classes, the posterior to prior odds B_{DI} may underestimate the evidence for dependent trait evolution. The alternative implementation of the Bayes factor may provide a more accurate estimate of support for dependent trait evolution. In this case, the posterior probability distribution sampled by the RJ-unconstrained chain, having excluded categories conforming to independent evolution, is compared to the posterior probability distribution sampled by the RJ-constrained chain (Section S1.1.3). At convergence, the RJ-constrained chain sampled states with mean $\log_e(\text{likelihood}) \pm \text{SD} = -20.12 \pm 0.79$ (range: -27.94 to -16.37 ; auto-correlation coefficient: $r = 0.003$; mean acceptance rate: 35.8%). Comparison of the $\log_e[H(\text{likelihood})]$ values yields $2\log_e(B_{DI}) \approx 2[(-20.56) - (-20.78)] = 0.44$, and thus no evidence for dependent evolution by the conservative criteria used (Table S2). This reflects the substantial overlap in the distributions of $\log_e(\text{likelihood})$ values for the states in the model of trait evolution sampled by the two chains, as shown in Figure 2 in the main text.

References

- Dyen, I., Kruskal, J. B., and Black, P. (1992). An Indoeuropean classification: a lexicostatistical experiment. *Transactions of the American Philosophical Society*, 82(5):1–132.
- Fortunato, L. (2009). *The evolution of kinship and marriage systems*. Doctor of philosophy in anthropology, University College London. Available from <http://discovery.ucl.ac.uk/18995/>.
- Fortunato, L. (2011a). Reconstructing the history of marriage strategies in Indo-European-speaking societies: monogamy and polygyny. *Human Biology*, 83(1):87–105.
- Fortunato, L. (2011b). Reconstructing the history of residence strategies in Indo-European-speaking societies: neo-, uxori-, and virilocality. *Human Biology*, 83(1):107–128.
- Fortunato, L. (2011c). Reconstructing the history of marriage and residence strategies in Indo-European-speaking societies. *Human Biology*, 83(1):129–135.
- Gray, J. P. (1999). A corrected *Ethnographic Atlas*. *World Cultures*, 10(1):24–136.
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R., and Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, 294(5550):2310–2314.
- Murdock, G. P. (1949). *Social structure*. The Free Press, New York, NY.
- Murdock, G. P. (1967). Ethnographic Atlas: a summary. *Ethnology*, 6(2):109–236.
- Pagel, M., Atkinson, Q. D., and Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, 449(7163):717–720.
- Pagel, M. and Meade, A. (2004). A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Systematic Biology*, 53(4):571–581.
- Pagel, M. and Meade, A. (2005). Bayesian estimation of correlated evolution across cultures: a case study of marriage systems and wealth transfer at marriage. In Mace, R., Holden, C. J., and Shennan, S., editors, *The evolution of cultural diversity: a phylogenetic approach*, chapter 13, pages 235–256. UCL Press, London.

- Pagel, M. and Meade, A. (2006). Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *The American Naturalist*, 167(6):808–825.
- Pagel, M. and Meade, A. (n.d.). BayesTraits *manual (version 1.0)*. School of Biological Sciences, University of Reading. Retrieved September 1, 2008, from <http://www.evolution.reading.ac.uk/BayesTraits.html>.
- Pagel, M., Meade, A., and Barker, D. (2004). Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology*, 53(5):673–684.
- Raftery, A. E. (1996). Hypothesis testing and model selection. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov chain Monte Carlo in practice*, chapter 10, pages 163–187. Chapman & Hall/CRC, London.