Human Biology

Volume 83 | Issue 1 Article 8

2011

Supporting Information: Reconstructing the History of Marriage and Residence Strategies in Indo-European-Speaking Societies

Laura Fortunato
Department of Anthropology, University College London, 14 Taviton Street, London WC1H 0BW, UK, fortunato@santafe.edu

Recommended Citation

Fortunato, Laura (2011) "Supporting Information: Reconstructing the History of Marriage and Residence Strategies in Indo-European-Speaking Societies," *Human Biology*: Vol. 83: Iss. 1, Article 8.
Available at: http://digitalcommons.wayne.edu/humbiol/vol83/iss1/8

Supporting Information: Reconstructing the History of Marriage and Residence Strategies in Indo-European-Speaking Societies

Abstract

This file provides additional information on the data and methods used in Fortunato (2011a,b), and discussion of the results of the fossilization of nodes Proto-Indo-Hittite (PIH) and Proto-Indo-European (PIE) for marriage and residence strategies.

Keywords

INDO-EUROPEAN, CULTURAL PHYLOGENETICS, MARRIAGE, MONOGAMY, POLYGYNY, AFFINAL TERMINOLOGY, RESIDENCE, NEOLOCALITY, UXORILOCALITY, VIRILOCALITY

Supporting Information

Reconstructing the History of Marriage and Residence Strategies in Indo-European—Speaking Societies

LAURA FORTUNATO[1]

This file provides additional information on the data and methods used in Fortunato (2011a,b), and discussion of the results of the fossilization of nodes Proto-Indo-Hittite (PIH) and Proto-Indo-European (PIE) for marriage and residence strategies.

Data and Methods

Below I provide details on the criteria used to collate the cross-cultural sample, with the cross-cultural data in table form, and information on the procedure used by Pagel et al. (2007) to infer the posterior probability distribution of trees on which I mapped the cross-cultural data. Finally, I provide a detailed description of the method used for the comparative analyses.

Cross-Cultural Data. Variable identifiers in this section follow Gray's (1999) *Ethnographic Atlas* (*EA*) codebook. I collated the cross-cultural sample by matching societies scored as speaking Indo-European (IE) languages (based on *EA* variable 98) with speech varieties in Dyen et al.'s (1992) IE basic vocabulary database, where needed using information from additional ethnographic and linguistic sources (e.g., Gordon 2005; Levinson 1991–1996; Price 1989; Ruhlen 1991). I also checked for correspondence between speech varieties in the linguistic database and the 62 societies in the *EA* with linguistic affiliation unknown and located in East Eurasia or in the Circum-Mediterranean region (based on *EA* variable 91).

In some cases, more than one speech variety in the linguistic database could be matched with the same society in the *EA*. For example, Dyen et al. (1992) include five entries for Greek: three for dialectal forms (Greek D, Greek K, Greek ML), one for modern Greek (Greek Mod), and one for modern spoken

---

[1] Department of Anthropology, University College London, 14 Taviton Street, London WC1H 0BW, UK. Current address: Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501.

Human Biology, February 2011, v. 83, no. 1, pp. 129–135

KEY WORDS: INDO-EUROPEAN, CULTURAL PHYLOGENETICS, MARRIAGE, MONOGAMY, POLYGYNY, AFFINAL TERMINOLOGY, RESIDENCE, NEOLOCALITY, UXORILOCALITY, VIRILOCALITY.

Greek (Greek MD), the latter compiled from dictionary data. In these cases, where available I selected the variety derived from dictionary data, which is likely to be less specific than other entries; alternatively, I selected the variety with data for the greatest number of meanings, or the first variety listed in Dyen et al. (1992, pp. 99–101).

The phylogenetic tree model used to represent how societies are related captures the process of diversification of taxa from a common ancestor; therefore, I included in the sample only societies located in Eurasia, corresponding to the geographic range of IE languages before 1492 CE (Diamond and Bellwood 2003). I excluded the Icelanders because the *EA* description for this society refers to 1100 CE, while the descriptions for the 27 societies included in the sample refer to the "ethnographic present," with dates ranging from 1880 to 1960 CE, and median 1945 CE (Murdock 1967). Table 1 includes the recoded data on marriage strategy and residence strategy (prevailing and alternative modes) for the 27 societies.

Tree Sample. Pagel et al. (2007) inferred the posterior probability distribution of trees from Dyen et al.'s (1992) IE basic vocabulary database, using the Bayesian Markov chain Monte Carlo (MCMC) phylogenetic tree-building method developed by Pagel and Meade (2004). The linguistic database includes word forms and cognacy judgments for 95 modern IE speech varieties (languages, dialects, and creoles) across the Swadesh 200-word list of items of basic vocabulary; two or more word forms are cognate if they share a common origin. Swadesh lists consist of cross-culturally universal items of vocabulary such as pronouns, body parts, and numerals, which are less prone to innovation and borrowing (i.e., horizontal transmission) than other meanings (Swadesh 1952).

The tree-building analysis was performed on a data matrix obtained from the linguistic database as follows. First, Pagel et al. (2007) excluded eleven speech varieties suspected of methodological bias by Dyen et al. (1992) and added data for three extinct varieties (Hittite, Tocharian A, Tocharian B) to be used as "outgroup" taxa. Outgroups provide information on the direction of change in the data by virtue of being distantly related to the groups under investigation, the "ingroup" taxa; they are used in tree-building for determining ancestor-descendant relationships (Felsenstein 2004, p. 6). As discussed in Fortunato (2011a), Hittite belongs to the extinct sister-group to the IE languages, the Anatolian sub-group; the two known dialects of Tocharian, A and B, are extinct IE speech varieties (Ruhlen 1991, p. 325). Second, Pagel et al. (2007) transformed the linguistic data into a binary matrix, with 87 rows corresponding to the speech varieties and 4049 columns corresponding to classes of cognates; speech varieties were coded for presence or absence of word forms belonging to each cognate class.

Comparative Analysis. For clarity, the discussion in this section focuses on the simplest case of one binary trait, as in the analysis in Fortunato (2011a); this is easily extended to the case of one ternary trait, as in the analysis in Fortunato

(2011b). The specifics of each case are detailed in the methods section of the papers. 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.).

Given a posterior probability sample of trees, BayesMultistate estimates the posterior probability distributions of rate parameters and of ancestral states through a Markov chain implementing the model of trait evolution. For one binary trait taking states 0 and 1, the model of evolution is defined by the rate parameters q_{01} and q_{10}, the likelihood function, the prior probability distributions of the rates, and the posterior probability distribution of trees in the tree sample.

Mathematically, a Markov chain is a random process in which the next state depends only on the current state, that is, the next state is independent of

Table 1. Recoded *EA* Data

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

a. Refers to the numbers in Figure 1 in Fortunato (2011a) and in Fortunato (2011b).

b. After Dyen et al. (1992).

c. After Gray (1999).

d. Recoded from *EA* variable 9; see Fortunato (2011a) for details.

e. Recoded from *EA* variable 12 for prevailing mode and 14 for alternative mode; see Fortunato (2011b) for details.

where the process was previously (Felsenstein 2004, p. 293). MCMC methods are a class of algorithms for sampling from a probability distribution, based on constructing a Markov chain that has as its stationary distribution the desired distribution (Huelsenbeck et al. 2001); for BayesMultistate, this is the posterior probability distribution of the parameters of interest to the comparative question. Specifically, BayesMultistate uses the Metropolis-Hastings algorithm (Hastings 1970; Metropolis et al. 1953) to simulate a random walk across the universe of possible states in the model of trait evolution and to periodically sample from the states visited. States in the chain correspond to a set of parameters values and a tree drawn from the tree sample. Schematically, starting from a random state, at each step in the chain a tree is drawn at random from the sample of trees and values are proposed for rate parameters. Whether a new state is accepted, and thus added to the sample, is determined by comparing the likelihood of the new state to the likelihood of the current state in the chain. States with better likelihood are always accepted, whereas states with worse likelihood are accepted only a proportion of the time equal to the ratio of the likelihood of the new state to the likelihood of the current state; if the new state is rejected, the current state is added to the sample again. Through this process, the chain visits states in proportion to their posterior probability; therefore, if the chain is run for a large number of steps, the distribution of states in the sample produced by the chain closely approximates their posterior probability distribution. Consequently, the proportion of the time that any state appears in the sample is a valid approximation of its posterior probability (Holder and Lewis 2003; Lewis 2001).

A particular implementation of MCMC methods, reversible jump (RJ) MCMC (Green 1995), can be used to additionally estimate the posterior probability distribution of the possible model categories. Four model categories are possible for one binary trait taking states 0 and 1, described by the rate parameters q_{01} and q_{10}: q_{01} and q_{10} may take distinct positive values, they may take the same positive value, or either one may be set to zero while the other takes a positive value. In this case, at each step in the chain q_{01} and q_{10} are assigned the same positive value, distinct positive values, or either one is set to zero while the other is assigned a positive value. Thus, an RJ-MCMC chain samples simultaneously from the posterior probability distributions of model categories and of the parameters in the model of trait evolution.

Combining estimates over the sample produced by the chain corresponds to “averaging” inferences over uncertainty in the phylogeny, in the parameters of the model of trait evolution, and, in the RJ-MCMC case, in the model itself. Crucially, the validity of the inferences depends on convergence of the chain to its stationary distribution, that is, to the posterior probability distribution of the parameters of interest to the comparative question; in turn, this depends on the ability of the chain to wander through “state space” effectively, which is determined by the chain specifications. The chain specifications used for each analysis are detailed in the methods sections of the papers.

Table 2. Summary Parameters for the Samples Returned by the Fossilization Chains for Marriage Strategy

	Fossil State	
Parameter	*M*	*P*
a. Node PIH		
Mean $\log_e$(likelihood) ± SD	−11.44 ± 0.78	−12.82 ± 0.80
Range of $\log_e$(likelihood) values	−20.91 to −9.91	−21.05 to −10.74
Autocorrelation coefficient r	0.001	−0.001
Mean acceptance rate	21.0%	24.8%
$\log_e[H(\text{likelihood})]$	−12.60	−13.43
b. Node PIE		
Mean $\log_e$(likelihood) ± SD	−11.27 ± 0.76	−13.19 ± 1.01
Range of $\log_e$(likelihood) values	−21.07 to −9.94	−21.98 to −10.74
Autocorrelation coefficient r	−0.001	0.001
Mean acceptance rate	21.3%	25.1%
$\log_e[H(\text{likelihood})]$	−12.38	−14.05

Results

In this section I discuss the fossilization of nodes PIH and PIE for marriage and residence strategies.

Fossilization of Nodes PIH and PIE for Marriage Strategy. At convergence, the chains fossilizing nodes PIH and PIE sampled states in the model of trait evolution as shown in Table 2.

Comparison of the $\log_e[H(\text{likelihood})]$ values yields $2\log_e(B_{\text{MP}}) \approx 1.66$ at node PIH (Table 2a), corresponding to no evidence for monogamy over polygyny by the conservative criteria used, and $2\log_e(B_{\text{MP}}) \approx 3.35$ at node PIE (Table 2b), corresponding to positive evidence for monogamy over polygyny.

Overall, these results confirm the ancestral state estimates obtained with nodes not fossilized; in particular, the magnitude of the evidence for monogamy at node PIE reflects the strong skew toward high values for state M in the posterior probability distribution (Figures 2 and 3 in Fortunato 2011a).

Fossilization of Nodes PIH and PIE for Residence Strategy. At convergence, the chains fossilizing nodes PIH and PIE sampled states in the model of trait evolution as shown in Table 3.

Comparison of the $\log_e[H(\text{likelihood})]$ values at node PIH yields $2\log_e(B_{\text{NU}}) \approx 0.15$, corresponding to no evidence for neo- over uxorilocality, $2\log_e(B_{\text{NV}}) \approx -1.36$, corresponding to no evidence for viri- over neolocality, and $2\log_e(B_{\text{UV}}) \approx -1.51$, corresponding to no evidence for viri- over uxorilocality (Table 3a).

Comparison of the $\log_e[H(\text{likelihood})]$ values at node PIE yields $2\log_e(B_{\text{NU}}) \approx 3.15$, corresponding to positive evidence for neo- over uxorilocality, $2\log_e(B_{\text{NV}}) \approx -4.36$, corresponding to positive evidence for viri- over

Table 3. Summary Parameters for the Samples Returned by the Fossilization Chains for Residence Strategy

Parameter	*Fossil State* *N*	*U*	*V*
a. Node PIH			
Mean $\log_e$(likelihood) ± SD	−49.68 ± 1.72	−49.51 ± 1.82	−49.62 ± 1.34
Range of $\log_e$(likelihood) values	−62.25 to −45.91	−61.97 to −44.04	−61.82 to −45.98
Autocorrelation coefficient r	0.025	0.083	0.026
Mean acceptance rate	26.2%	24.7%	29.2%
$\log_e[H(\text{likelihood})]$	−52.64	−52.72	−51.96
b. Node PIE			
Mean $\log_e$(likelihood) ± SD	−50.95 ± 1.89	−52.07 ± 2.34	−50.34 ± 1.51
Range of $\log_e$(likelihood) values	−64.81 to −46.26	−66.43 to −46.00	−61.60 to −45.85
Autocorrelation coefficient r	0.108	0.242	0.028
Mean acceptance rate	26.1%	16.6%	27.6%
$\log_e[H(\text{likelihood})]$	−54.70	−56.28	−52.52

neolocality, and $2\log_e(B_{\mathrm{UV}}) \approx -7.51$, corresponding to strong evidence for viri- over uxorilocality (Table 3b).

As noted in Fortunato (2011b), these posterior probability distributions effectively “average” the fit of the fossil state at a node over the three sites, that is, over prevailing and alternative modes of residence; therefore, the strength of the evidence in favor of any particular state at the node is likely to be underestimated. This may partly explain the lack of evidence for any state, at least by the conservative criteria used, at node PIH. In any case, the results broadly confirm the ancestral state estimates obtained with nodes not fossilized. The lack of evidence for any state at node PIH may partly reflect the moderate skew in the posterior probability distributions of the three states at this node, for the two modes of residence (Figure 2 in Fortunato 2011b). By contrast, the magnitude of the evidence for virilocality at node PIE reflects the strong skew in the posterior probability distribution toward high values for state V as prevailing mode of residence (Figure 3 in Fortunato 2011b).

Received 28 April 2010; revision accepted for publication 16 September 2010.

Literature Cited

Diamond, J., and P. Bellwood. 2003. Farmers and their languages: The first expansions. *Science* 300:597–603.

Dyen, I., J. B. Kruskal, and P. Black. 1992. An Indoeuropean classification: A lexicostatistical experiment. *Trans. Am. Philos. Soc.* 82:1–132.

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

Fortunato, L. 2011a. Reconstructing the history of marriage strategies in Indo-European—speaking societies: Monogamy and polygyny. *Hum. Biol.* 83(1):87–105.

Fortunato, L. 2011b. Reconstructing the history of residence strategies in Indo-European—speaking societies: Neo-, uxori-, and virilocality. *Hum. Biol.* 83(1):107–128.

Gordon, R. G., ed. 2005. *Ethnologue: Languages of the World*, 15th ed. Dallas, TX: SIL International. http://www.ethnologue.com.

Gray, J. P. 1999. A corrected *Ethnographic Atlas. World Cultures* 10:24–136.

Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika.* 82:711–732.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika.* 57:97–109.

Holder, M., and P. O. Lewis. 2003. Phylogeny estimation: Traditional and Bayesian approaches. *Nat. Rev. Genet.* 4:275–284.

Huelsenbeck, J. P., F. Ronquist, R. Nielsen et al. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–2314.

Levinson, D., ed. 1991–1996. *Encyclopedia of World Cultures.* Boston, MA: G. K. Hall & Co.

Lewis, P. O. 2001. Phylogenetic systematics turns over a new leaf. *Trends Ecol. Evol.* 16:30–37.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth et al. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092.

Murdock, G. P. 1967. Ethnographic Atlas: A summary. *Ethnology.* 6:109–236.

Pagel, M., Q. D. Atkinson, and A. Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature.* 449:717–720.

Pagel, M., and A. Meade. 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.* 53:571–581.

Pagel, M., and A. Meade. 2005. Bayesian estimation of correlated evolution across cultures: A case study of marriage systems and wealth transfer at marriage. In *The Evolution of Cultural Diversity: A Phylogenetic Approach*, R. Mace, C. J. Holden, and S. Shennan, eds. London, UK: UCL Press, 235–256.

Pagel, M., and A. Meade. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* 167:808–825.

Pagel, M., and A. Meade. n.d. BayesTraits *manual (version 1.0).* School of Biological Sciences, University of Reading. Retrieved September 1, 2008. http://www.evolution.reading.ac.uk/BayesTraits.html.

Pagel, M., A. Meade, and D. Barker. 2004. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* 53:673–684.

Price, D. H. 1989. *Atlas of World Cultures: A Geographical Guide to Ethnographic Literature.* Newbury Park, CA: Sage.

Ruhlen, M. 1991. *A Guide to the World's Languages*, vol. 1: *Classification.* London: Edward Arnold.

Swadesh, M. 1952. Lexico-statistic dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proc. Am. Philos. Soc.* 96:452–463.