Your place or mine? A phylogenetic comparative analysis of marital residence in Indo-European and Austronesian societies

Laura Fortunato[1,2,*,†] and Fiona Jordan[1,2,‡]

[1]*Department of Anthropology, University College London, 14 Taviton Street, London WC1H 0BW, UK*
[2]*AHRC Centre for the Evolution of Cultural Diversity, Institute of Archaeology, University College London, 31–34 Gordon Square, London WC1H 0PY, UK*

Accurate reconstruction of prehistoric social organization is important if we are to put together satisfactory multidisciplinary scenarios about, for example, the dispersal of human groups. Such considerations apply in the case of Indo-European and Austronesian, two large-scale language families that are thought to represent Neolithic expansions. Ancestral kinship patterns have mostly been inferred through reconstruction of kin terminologies in ancestral proto-languages using the linguistic comparative method, and through geographical or distributional arguments based on the comparative patterns of kin terms and ethnographic kinship 'facts'. While these approaches are detailed and valuable, the processes through which conclusions have been drawn from the data fail to provide explicit criteria for systematic testing of alternative hypotheses. Here, we use language trees derived using phylogenetic tree-building techniques on Indo-European and Austronesian vocabulary data. With these trees, ethnographic data and Bayesian phylogenetic comparative methods, we statistically reconstruct past marital residence and infer rates of cultural change between different residence forms, showing Proto-Indo-European to be virilocal and Proto-Malayo-Polynesian uxorilocal. The instability of uxorilocality and the rare loss of virilocality once gained emerge as common features of both families.

Keywords: phylogenetic comparative methods; cultural phylogenetics; post-marital residence; Indo-European; Austronesian; human social organization

1. INTRODUCTION

Marital residence norms are an important determinant of human kinship organization. By regulating the movement of people, these norms shape the pattern of genetic variation within and across populations. Knowledge of this feature of social organization is therefore crucial to our understanding of human demographic history (e.g. Seielstad *et al.* 1998; Wilkins & Marlowe 2006). Until recently, our ability to specify the behavioural strategies of people in prehistory was speculative at best. Ancestral kinship patterns have been inferred (i) through reconstruction of kin terminologies in ancestral proto-languages using the linguistic comparative method and (ii) through geographical or distributional arguments based on the comparative patterns of kin terms and ethnographic kinship 'facts' (e.g. Murdock 1949; Blust 1980; Mallory 1997). We stress that these approaches are

detailed and carefully described, and valuable to the study of prehistory. However, the processes through which conclusions have been drawn from the data fail to provide explicit criteria for systematic testing of alternative hypotheses. Accurate reconstruction of prehistoric social organization is important if we are to put together satisfactory multidisciplinary scenarios about, for example, the dispersal of human groups. Such considerations apply in the case of Indo-European (IE) and Austronesian (AN), two large-scale language families that are thought to be Neolithic expansions associated with new domestication technologies (Diamond & Bellwood 2003). In this paper, we discuss the results of phylogenetic comparative analyses of marital residence in societies speaking IE and AN languages, and show how this approach progresses speculation into a testable scientific framework.

Work in both language families illustrates the importance of correctly inferring ancestral social organization. For example, Gimbutas (1991) excluded Anatolia and southeastern and central Europe as potential homelands of the IE language family on the grounds that the Neolithic societies of these regions were 'matrifocal', while early IE society was reconstructed as practising virilocality (i.e. residence of married couples with or near the husband's kin) and patrilineality (Mallory 1997). However, interpretations of the linguistic evidence seem to be strongly biased towards virilocality by the purportedly

* Author for correspondence (l.fortunato@ucl.ac.uk).
† Present address: Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM87501, USA.
‡ Present address: Evolutionary Processes in Language and Culture, Max Planck Institute for Psycholinguistics, PB 310, Nijmegen, The Netherlands.

'male-centred' structure of early IE society (Clackson 2007; e.g. Anthony 2007). The prevalence of virilocality among the historically attested IE societies is used to bolster these interpretations (e.g. Mallory 1997), but the ethnographic evidence can also be used to support alternative scenarios. For example, based on cross-cultural variation in social systems, Murdock (1949, p. 349) reconstructed 'an Eskimo type of social structure in the prehistory of the Indo-European peoples', which is characterized in its typical form by monogamous marriage, independent nuclear families and neolocal residence (Murdock 1949).

While diverse in terms of marital residence norms, AN societies are noted for their flexibility (Lane 1961) and their 'matricentric orientation', that is, a theme of uxorilocality (i.e. residence of married couples with or near the wife's kin) and matrilineality (Burton *et al.* 1996). Pacific scholars have debated the nature of early AN social organization for many years with little apparent consensus (Van Wouden 1935 [1968]; Murdock 1949; Blust 1980). As with IE, inferences have relied on linguistic reconstructions and inferences from comparative ethnography. More recently, Hage (1998) and Marck (Hage & Marck 2003; Marck 2008) hypothesized that uxorilocality characterized ancestral Oceanic society (the branch of the family including Polynesian and other Remote Oceanic societies). A matri-biased social organization in ancestral Oceanic peoples would therefore have restricted female genetic diversity while increasing male diversity as non-AN men married in. Uxorilocality is thus consistent with the divergent mtDNA and Y-chromosome patterns seen in Pacific human genetics, and some geneticists are beginning to work within this paradigm (e.g. Kayser *et al.* 2008; for a review see Hurles *et al.* 2003).

These examples illustrate how putative ancestral kinship patterns are invoked to constrain hypotheses (as in IE) or to explain conflicting evidence (as in AN) about the past. Resolving questions about past social structure will thus play a large part in correctly describing population prehistory. Phylogenetic comparative methods offer a rigorous statistical framework for reconstructing the pattern of change in cultural traits, and provide insights into features of social organization that are not preserved in the archaeological or historical records (Mace & Pagel 1994). Even within a small literature, phylogenetic comparative analyses of human cultural traits have concentrated on aspects of kinship (Fortunato 2008). The fact that kinship systems are organized in a restricted set of all the combinatorial possibilities available (e.g. Nerlove & Romney 1967) suggests that selective forces are at work to optimize these features of human social behaviour. Because they involve a co-dependent set of individuals, at the group level, kinship norms are likely to be stable for many generations, particularly if they represent effective behavioural strategies (e.g. Guglielmino *et al.* 1995). It then stands that if human societies have shared trajectories of cultural evolution, kinship is a likely locus for us to discover such commonalities. Here, we focus our discussion on comparison of the inferred patterns of change in marital residence across the two ethno-

linguistic groups; the reconstructions for the two ethno-linguistic groups are discussed in detail in Fortunato (in press) for IE and in Jordan *et al.* (2009) for AN. Following Richards (1950), who noted that matrilineal systems involve tensions between male authority and the female focus of kinship relations, we suggest that one predicted commonality might be higher rates of change away from uxorilocal residence to other forms.

2. THE PHYLOGENETIC COMPARATIVE APPROACH

Phylogenetic comparative methods work by reconstructing evolutionary pathways that are likely to have produced the observed distribution of traits of interest across a sample of taxa. This requires a phylogenetic tree representing the historical relationships among the taxa, and a model of how the traits have evolved on the tree (Felsenstein 1985; Harvey & Pagel 1991). In a cross-cultural framework, the model of trait evolution is inferred statistically from ethnographic comparative data mapped onto phylogenetic trees derived from genetic or linguistic data (Mace & Pagel 1994; e.g. Holden & Mace 1997, 2003; Fortunato *et al.* 2006).

Here, we reconstruct ancestral states of marital residence across samples of societies speaking IE and AN languages, using comparative data from ethnographic sources and phylogenetic trees derived from linguistic data. We use phylogenetic comparative methods in a Bayesian Markov chain Monte Carlo (MCMC) framework (Pagel *et al.* 2004; Pagel & Meade 2005, 2006) to estimate the posterior probability distributions of parameters of interest to the comparative question (e.g. ancestral state probabilities at internal nodes, rates of trait change). 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 (Huelsenbeck *et al.* 2001; Lewis 2001). Because posterior probabilities cannot feasibly be computed analytically, posterior probability distributions are inferred instead using an MCMC sampling algorithm. This distributional approach provides information about the degree of statistical uncertainty in the cultural trait reconstructions. Relatedly, this approach makes it possible to account for the effect of uncertainty in the phylogenetic tree model representing population history, a non-trivial consideration in the study of cultural traits as a single branching tree is unlikely to accurately represent human population history (Boyd *et al.* 1997): the estimation of parameters over a probability sample of trees yields estimates that are not dependent on any specific phylogenetic hypothesis. Finally, parameters can be estimated over different models of trait evolution, and this yields estimates that are not dependent on any specific model of how the cultural traits have evolved.

(a) *Tree samples*
We used the posterior probability samples of language trees published in Pagel *et al.* (2007) for IE and in Jordan *et al.* (2009) for AN. These samples were

themselves obtained from phylogenetic tree-building analyses of binary matrices showing the presence/absence of cognate terms from the 200-word list of basic vocabulary for 87 IE languages (data from Dyen *et al.* 1992) and the 210-word list for 400 AN languages (data from Greenhill *et al.* 2008; Gray *et al.* 2009), using the Bayesian MCMC method implemented in BAYESPHYLOGENIES (Pagel & Meade 2004). This method generates a sample of phylogenetic trees in which trees appear in proportion to their posterior probability. The samples included 750 trees for IE, and 1000 for AN; the size of the tree sample is arbitrarily large, and is determined by the specifications (e.g. length of the MCMC chain and sampling period) of the tree-building analysis. We pruned the parent trees in the samples to retain only those taxa for which we had corresponding cultural data in the ethnographic sources used ($n = 27$ plus outgroup Hittite for IE and $n = 135$ for AN); the criteria used for matching the linguistic and cultural taxa are outlined in §2b.

(i) *Outgroups and proto-societies*

Outgroup taxa are used in tree-building analyses for determining ancestor–descendant relationships; they provide information on the direction of change in the data (in this case, in the linguistic data), by virtue of being distantly related to the taxa under investigation, the ingroup taxa. The tree-building analysis for IE used Hittite, Tocharian A and Tocharian B as outgroups (Pagel *et al.* 2007). Hittite belongs to the extinct sister group of the IE languages, the Anatolian clade; together, the Anatolian and IE clades form the Indo-Hittite language family. The two known dialects of Tocharian, A and B, are speech varieties representing an extinct IE clade (Ruhlen 1991). We use the term 'Proto-Indo-Hittite' for the hypothetical ancestor of Indo-Hittite languages, and 'Proto-Indo-European' (PIE) for the hypothetical ancestor of IE languages, and for the hypothetical 'proto-societies' that spoke them. For consistency with previous work (Fortunato *et al.* 2006; Fortunato & Mace 2009), Hittite was retained in the tree sample, but was assigned no marriage strategy data for the purpose of the comparative analysis (§2b). In AN, the languages used as outgroup taxa in the tree-building analysis (Gray *et al.* 2009) were not retained in the tree sample as no corresponding cultural data could be found. Thus, the root of the tree corresponds to 'Proto-Austronesian' (PAN), whereas 'Proto-Malayo-Polynesian' corresponds to the hypothetical ancestor of all non-Formosan AN languages (Gray *et al.* 2009). For each set of analyses, we present a consensus tree summarizing the tree sample, but the comparative analyses were performed over the two tree samples in their entirety.

(b) *Coding residence data*

We matched languages to ethnographic data on marital residence using the geographical and descriptive information on societies in the anthropological literature. The IE analyses used data from Murdock's (1967) *Ethnographic Atlas*, based on the updated electronic version by Gray (1999). Only societies located in Eurasia were included in the sample, corresponding

to the geographical range of IE languages before 1492 CE (Diamond & Bellwood 2003). In addition to the *Ethnographic Atlas*, the AN analyses used data from ethnographic encyclopaedias (LeBar 1975; Levinson 1990) and relevant ethnographic literature or fieldworkers (Jordan *et al.* 2009). The *Ethnographic Atlas* scores societies separately for prevailing and alternative modes of marital residence, the latter defined as 'culturally patterned alternatives to, or significant deviations from, the prevailing profile' (Murdock 1967, p. 48). For AN, the data from additional sources were coded consistently with the *Ethnographic Atlas*.

In order to give higher weight to the prevailing mode of residence, each society was assigned three columns of data: two identical columns specifying the prevailing pattern, and a third column specifying the alternative pattern; the prevailing mode was used at all three columns for societies scored as not presenting an alternative mode. For both the prevailing and alternative patterns, we coded societies as practising neolocality (i.e. residence apart from the kin of either spouse; state N), uxorilocality (i.e. residence with or near the wife's kin; state U) or virilocality (i.e. residence with or near the husband's kin; state V). Ambilocal societies, where married couples take residence optionally with (or near) the kin of either spouse, and with approximately equal frequency, were assigned the dual state UV. Consistently, in the comparative analysis, these societies are treated as taking either state with equal probability (§2c). Missing information was coded as such (§2c). Below, we discuss results focusing on the changes in the prevailing residence pattern across the two tree samples.

(c) *Estimation of ancestral states*

We used the phylogenetic comparative method implemented in BAYESMULTISTATE, available as part of the BAYESTRAITS package from http://www.evolution.rdg.ac.uk/BayesTraits (Pagel *et al.* 2004; Pagel & Meade 2005, 2006). Given the comparative data and tree sample, BAYESMULTISTATE uses a continuous-time Markov model to describe the evolution of the trait of interest along the branches of a phylogeny. Under this model, the trait 'residence' can switch repeatedly between its three states, N, U and V, in any of the branches of a tree. In analyses with multiple sites—in this case, the three columns of data specifying the prevailing and alternative modes of residence—BAYESMULTISTATE uses information from the sites simultaneously to estimate a single set of rate parameters specifying the model of trait evolution. Three states require six rate parameters specifying the possible transitions—in this case, q_{NU}, q_{NV}, q_{UN}, q_{UV}, q_{VN} and q_{VU}. These parameters measure the instantaneous rates of change from one state to another, and are used to define the probabilities of these changes, the character states at internal nodes on a tree and the likelihood of the data (Pagel 1994, 1997, 1999). In the likelihood calculations, BAYESMULTISTATE treats taxa that are assigned multiple states, like the ambilocal societies (§2b), as taking those states with equal probability at the relevant site; similarly, it treats taxa with missing data as taking any state with equal probability.

The Bayesian MCMC implementation of BAYES-MULTISTATE estimates the posterior probability distributions of rate parameters and ancestral character states (Pagel *et al.* 2004; Pagel & Meade 2005). All analyses used the program in reversible-jump mode, which additionally estimates the posterior probability distribution of the possible models of trait evolution specified by the six rate parameters (Pagel & Meade 2006). The reversible-jump procedure outputs a model string describing the rate parameters such that rates are assigned to classes denoted by ordered integers or a 'zero bin' depicted by Z. For example, the model string 00011Z assigns q_{NU}, q_{NV} and q_{UN} all to an internally equivalent rate class (0) that is slower than class (1) to which q_{UV} and q_{VN} are assigned. Rate q_{VU} is set to zero.

The means of the posterior probability distribution of ancestral states at internal nodes on the consensus tree are combined with the posterior probability of each node, which represents the probability that the node exists (Lewis 2001), and is denoted as $p(\text{node})$. For example, for a given node, BAYESMULTISTATE may return a posterior probability distribution with a mean of $0.8 \pm$ s.d. for virilocality; this is denoted $p(\text{V}|\text{node}) \pm$ s.d. If the node is present in all trees, i.e. $p(\text{node}) = 1.00$, we accept the 0.8 value as the posterior probability of virilocality at that node. However, if the node is only present in 60 per cent of the trees, i.e. $p(\text{node}) = 0.60$, we report the 'combined probability' for virilocality, $p(\text{V}) = p(\text{V}|\text{node}) \times p(\text{node}) = 0.8 \times 0.6 = 0.48$. A value of 0.7 for the combined probabilities represents an acceptable value of certainty for an ancestral state at a node (M. Pagel 2006, personal communication).

The MCMC chain specifications were determined by examining the results of preliminary maximum-likelihood and MCMC runs. Different numbers of taxa and the probabilistic nature of the tree samples meant that the same specifications could not be applied to both datasets; however, where possible, we used comparable values (see electronic supplementary material, table S1, for details).

(d) *Testing*
We tested the ancestral state reconstructions at the root (Proto-Indo-Hittite for IE and PAN for AN) and a historically significant basal node in each family (PIE for IE and Proto-Malayo-Polynesian for AN). These tests 'fixed' each node to be one of the three possible states (N, V and U), in turn. BAYESMULTISTATE does not allow sites to be fossilized separately; therefore, each run fixed all three sites to the same state. We determined which fossilized state had relatively higher support at a given node using Bayes factors. Following Pagel & Meade (2006), we took twice the difference in the logarithm of the harmonic mean of the likelihoods for pairs of runs; the resulting values represent a summary of the evidence for one state over another at a given node. Based on Raftery's (1996) logarithmic scale for interpretation of the Bayes factors, values between 0 and 2 are barely worth mentioning, values between 2 and 5 represent positive evidence, values between 5 and 10 strong evidence and values greater than 10 very strong evidence.

3. RESULTS
Figures 1 and 2 present the consensus trees summarizing the two phylogenies. The IE phylogeny in figure 1 is shown with all nodes and reconstructions. The AN phylogeny (figure 2) is depicted in condensed form with clades collapsed; the size of a clade is proportional to the number of daughter societies (see electronic supplementary material, table S2, for a listing of societies within each clade). Clade triangles terminate in a bar shaded proportionally to reflect the frequency of residence patterns within each clade. Below, we report the major findings of ancestral state reconstruction for the two families, followed by a comparative discussion on the relative rates of change in residence.

(a) *Indo-European*
For the prevailing mode of residence, nodes Proto-Indo-Hittite and PIE reconstructed as virilocal with posterior probabilities of $p(\text{V}|\text{node}) = p(\text{V}) = 0.64 \pm 0.14$ and $p(\text{V}|\text{node}) = p(\text{V}) = 0.90 \pm 0.12$, respectively (figure 3). Virilocality reconstructed with high posterior probabilities through to nodes A, B and C (in all cases $p(\text{V}|\text{node}) \geq 0.85$), but the confidence that can be placed in these inferences is limited by the degree of phylogenetic uncertainty at these nodes. Uncertainty in the reconstructions at the base of the tree means that a host of scenarios can explain the observed distribution of states at the tips. Node D (the common ancestor of societies speaking Italic, Germanic and Celtic languages) yielded posterior probabilities of $p(\text{V}|\text{node}) = 0.40 \pm 0.10$ for virilocality and $p(\text{N}|\text{node}) = 0.40 \pm 0.15$ for neolocality; additionally, this node is found in only 78 per cent of trees in the sample, i.e. $p(\text{node}) = 0.78$. However, neolocality reconstructed with high posterior probabilities within the Italic clade. Nodes E (the common ancestor of societies speaking Indian and Iranian languages) and F (the common ancestor of societies speaking Baltic and Slavic languages) reconstructed as virilocal with posterior probabilities of $p(\text{V}|\text{node}) = p(\text{V}) = 0.87 \pm 0.10$ and $p(\text{V}|\text{node}) = p(\text{V}) = 0.92 \pm 0.08$, respectively. Virilocality reconstructed with high posterior probabilities within the Indo-Iranian and Balto-Slavic clades.

In agreement with the ancestral state reconstructions, at node PIE the fossilization analyses returned strong evidence for virilocality over uxorilocality (Bayes factor 7.51), positive evidence for virilocality over neolocality (Bayes factor 4.36) and positive evidence for neolocality over uxorilocality (Bayes factor 3.15). The pattern was weaker at node Proto-Indo-Hittite, with values of the Bayes factor less than 2 in all cases.

(b) *Austronesian*
Uxorilocality is securely reconstructed for Proto-Malayo-Polynesian ($p(\text{U}|\text{node}) = p(\text{U}) = 0.96 \pm 0.06$), and many daughter subgroups and societies in the Island Southeast Asian region (e.g. Proto-Philippines, Sumatran societies) still retain this pattern. The root
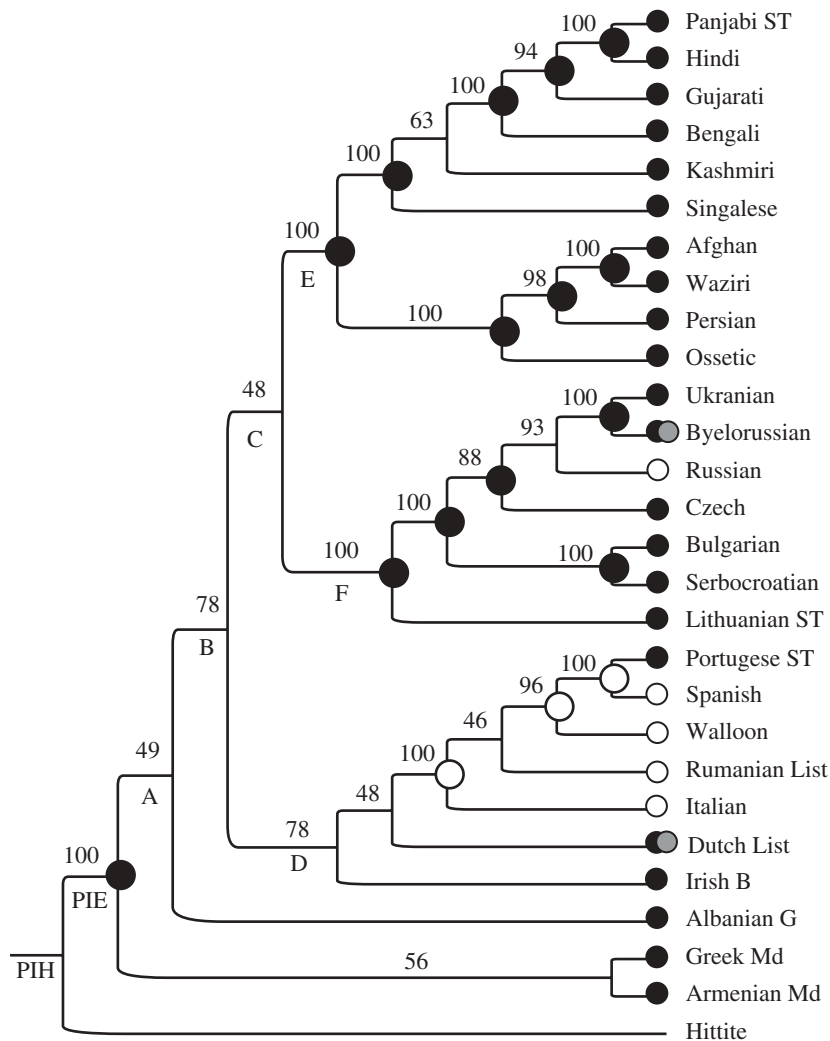
Figure 1. Fifty per cent majority rule consensus phylogeny (including compatible groupings) summarizing the sample of 750 trees for 27 IE languages plus the outgroup Hittite. The colour of the dots at the tips depicts a society's prevailing mode of marital residence: white, neolocality; grey, uxorilocality; black, virilocality; black and grey, ambilocality. The value above each node is the node's posterior probability, as a percentage. The dots at the nodes indicate the ancestral states of residence (white, neolocality; grey, uxorilocality; black, virilocality; black and grey, ambilocality). for the node (black $p(V) \geq 0.70$, grey $p(N) \geq 0.70$); nodes with no dots have combined probability less than 0.70 for all states. PIH, Proto-Indo-Hittite; PIE, Proto-Indo-European.

PAN cannot be reconstructed with certainty in this three-state analysis that considers both prevailing and alternate residence forms, in comparison to Jordan *et al.* (2009) who considered only the prevailing mode of residence and two states (uxorilocal/virilocal). Here, for prevailing mode of residence, PAN is uxorilocal ($p(U|\text{node}) = p(U) = 0.42 \pm 0.10$) more than virilocal ($p(V|\text{node}) = p(V) = 0.38 \pm 0.09$) or neolocal ($p(N|\text{node}) = p(N) = 0.20 \pm 0.10$), but there is considerable overlap in the distributions of these probabilities, leaving PAN an ambiguously reconstructed node (figure 4a). Alternative mode of residence at PAN is more securely uxorilocal ($p(U|\text{node}) = p(U) = 0.67 \pm 0.13$). Early AN societies thus do have a bias towards uxorilocality, suggesting that virilocality was a later development in the AN family as a whole. Switches to prevailing virilocality occur in many societies surrounding the island of New Guinea (clades such as Oceanic 1–3, South Halmahera–West New Guinea and Central Malayo-Polynesian 1 and 2), though some retain uxorilocality, especially as an

alternative strategy. The well-defined Proto-Oceanic node is present in all trees and reconstructs with prevailing virilocality ($p(V|\text{node}) = p(V) = 0.72 \pm 0.18$) and alternative uxorilocality ($p(U|\text{node}) = p(U) = 0.77 \pm 0.16$). No ancestral nodes robustly reconstruct as neolocal, and over all trees $p(N|\text{node})$ is rarely greater than 0.33. When tested using the fossilization procedure, there is positive evidence in favour of PAN uxorilocality over virilocality (Bayes factor 3.88) and neolocality (Bayes factor 3.44), and very strong evidence for Proto-Malayo-Polynesian uxorilocality over virilocality (Bayes factor 16.42) and neolocality (Bayes factor 15.04).

(c) *Rates of change*

Table 1 presents the mean values of the rate parameters over all model categories sampled by the chains; because we cannot directly compare parameter values across the two ethnolinguistic groups, we have scaled them against $q_{VU} = 1$. For each group, the
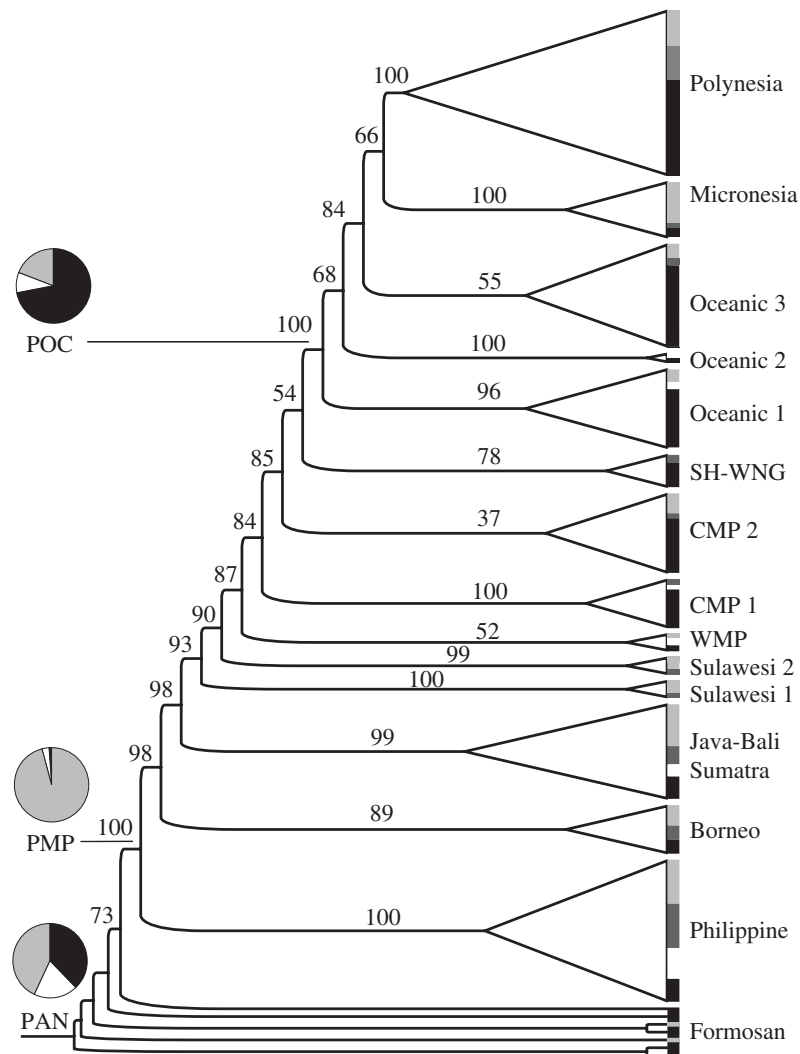
Figure 2. Fifty per cent majority rule consensus phylogeny (including compatible groupings) summarizing the sample of 1000 trees for 135 Austronesian languages. Collapsed clades are proportional in size to number of taxa, and terminate in a bar shaded proportional to residence patterns within that clade (white, neolocality; light grey, uxorilocality; black, virilocality; dark grey, ambilocality). The value above each node is the node's posterior probability, as a percentage. Three major nodes only (Proto-Austronesian, PAN; Proto-Malayo-Polynesian, PMP; Proto-Oceanic, POC) are shown shaded according to the ancestral state reconstruction. CMP, Central Malayo-Polynesian; SHWNG, South Halmahera–West New Guinea; WMP, Western Malayo-Polynesian.

table also includes the string representing the model category sampled most frequently by the chain.

In IE, the top reversible-jump model category accounts for 17 per cent of all points sampled by the chain. Here, rate parameter q_{VU} was always assigned to the zero bin; in fact, it was assigned to the zero bin in the top 81 per cent of points sampled by the chain. The mean value of q_{VU} over all sampled points was an order of magnitude smaller than the mean values of the other five rate parameters (0.2 compared with mean values ranging from 3.9 for q_{VN} to 8.5 for q_{NV}). The distribution of states at the tips of the tree and the reconstructed ancestral states indicate that transitions from viri- to uxorilocality in the prevailing mode of residence are rare, occurring only in the branch leading to Byelorussian and possibly in the branch leading to Dutch (figure 1); only three societies (Armenian, Italian and Singhalese) practise alternative uxorilocality. This means that the

acquisition of uxorilocality is more likely to have occurred through neolocality than through virilocality throughout the history of IE-speaking societies.

In AN, the top reversible-jump model category accounts for 23 per cent of all sampled points and captures the overall dynamics of how residence changes in these societies. In this, and in the first 94 per cent of sampled points, q_{VN} is assigned to the zero bin. The value of q_{VN} over all sampled points is effectively zero (0.6 compared with mean values greater than 27.6): thus, changes from viri- to neolocality are rare. Further, rates from viri- to uxorilocality and from uxori- to neolocality are always assigned to the slow rate category. Other transitions vary equally between the slower and faster rates, reflected in their mean values. The general pattern emerges that in AN residence, changes towards neolocality are uncommon, and transitions towards virilocality happen frequently. Uxorilocal societies are 1.5 times more likely to switch to virilocality ($q_{UV} = 40.4$)
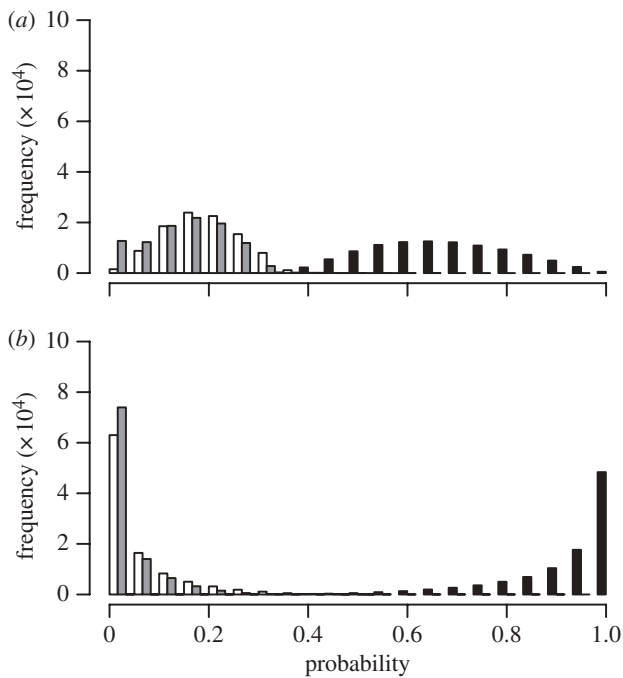
Figure 3. Posterior probability distributions of reconstructed ancestral states for nodes corresponding to (*a*) Proto-Indo-Hittite and (*b*) PIE. Bar colours match the residence codings in figure 1: white, neolocality; grey, uxorilocality; black, virilocality.
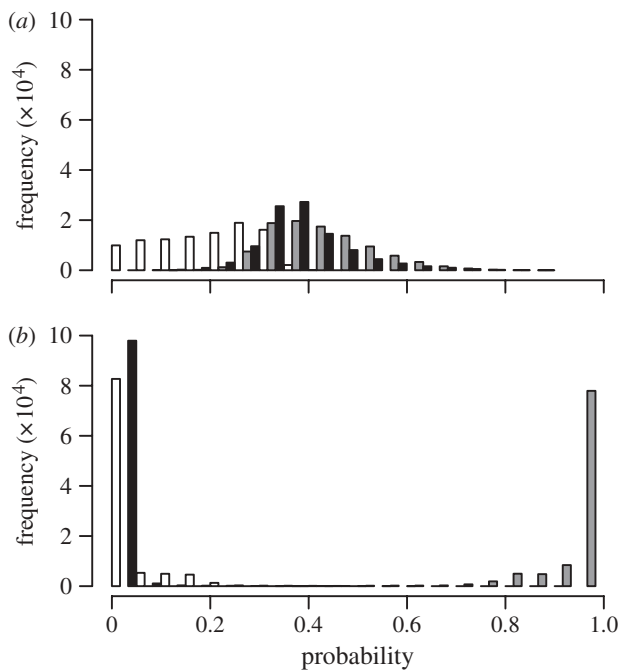

Figure 4. Posterior probability distributions of reconstructed ancestral states for nodes corresponding to (*a*) PAN and (*b*) Proto-Malayo-Polynesian. White, neolocality; grey, uxorilocality; black, virilocality.

than virilocal societies are to switch to uxorilocality ($q_{VU} = 28.6$); additionally, there are no instances where $q_{VU} > q_{UV}$.

Overall, the analyses suggest that, in both ethnolinguistic families, the dynamics of evolutionary change in the residence strategy can be described by a model of trait evolution based on a small number of non-

zero rate classes (the mean number of non-zero rate parameters is 1.76 for IE and 2.04 for AN). More generally, the results suggest that in both IE and AN the loss of virilocality is a rare event, as indicated by the relative values of the rate parameters capturing these transitions (q_{VU} and q_{VN}). It is especially the case that changes from uxori- to virilocality (specified by parameter q_{UV}) occur at a higher rate than the reverse transition (specified by parameter q_{VU}): q_{UV} is over 30 times more likely than q_{VU} in IE, and one and a half times more likely in AN (table 1).

4. DISCUSSION

Using phylogenetic comparative methods and ethnolinguistic information on two large cultural families, we have reconstructed an important aspect of the social structure of peoples who lived over 5000 years ago.

The reconstruction of early IE virilocality is in line with the prevalent scenario derived from the linguistic evidence (Mallory 1997); as noted above, however, reconstructions of virilocality based on the linguistic evidence are plagued by substantial bias in interpretation, and several alternatives are at least equally plausible (Clackson 2007). The uncertainty in the reconstruction for Proto-Indo-Hittite reflects disagreements in the literature about the earliest residence pattern of IE peoples (Clackson 2007) and suggests that, for this point in time, we can place limited confidence in inferences about this aspect of social organization drawn from cross-cultural data. The reconstruction of early IE virilocality concurs with recent archaeological evidence based on strontium isotope analyses of Neolithic burials in Germany, which indicate the migration of females in adulthood (Price *et al.* 2001; Bentley *et al.* 2002; Haak *et al.* 2008; see discussion in Fortunato in press).

In AN, early uxorilocality appears to be robustly supported in Proto-Malayo-Polynesian and as an alternate option in PAN. This is in line with some interpretations of PAN and Proto-Malayo-Polynesian kinship terminologies (Blust 1980), but, as with IE, here we provide independent confirmation from cross-cultural data. More recent work attempting to reconcile the different patterns of uniparental genetic markers seen in the Pacific has suggested that uxorilocality was a later development in AN, i.e. in Proto-Oceanic (Hage 1998; Hage & Marck 2003; Kayser *et al.* 2008). However, our findings suggest that this period of uxorilocality was earlier in time; our comparative methods may not be able to reconstruct this form of residence for Proto-Oceanic because many daughter societies have, while retaining an uxorilocal option, since switched to virilocality as the prevailing mode perhaps because of cultural contact with nearby non-AN ('Papuan') societies (Jordan *et al.* 2009). Further work is required to identify independent 'markers' for contact that might allow us to systematically address hypotheses about cultural borrowings.

The inferred model of trait evolution shows that in both IE and AN changes from uxori- to virilocality occur at a higher rate than the reverse transition. This may reflect the instability of 'matricentric'

Table 1. Summary of the model of trait evolution for residence strategy. N, neolocality; U, uxorilocality; V, virilocality.

	analysis	rate parameters					
		q_{NU}	q_{NV}	q_{UN}	q_{UV}	q_{VN}	q_{VU}
mean value[a]	IE	5.9	8.5	8.0	6.4	3.9	0.2
	AN	43	83	27.6	40.4	0.6	28.6
scaled value[a]	IE	29.5	42.5	40	32	19.5	1
	AN	1.5	2.9	0.9	1.4	0.02	1
rate class[b]	IE	0	0	0	0	0	Z
	AN	0	1	0	0	Z	0

[a]Mean and scaled values over all model categories sampled by the chain. The mean values are scaled by setting $q_{VU} = 1$.
[b]Rate class to which the rate parameter was assigned in the model sampled most frequently by the chain. 'Z' denotes rate parameters assigned to the zero bin (see §2*c*).

systems (e.g. systems involving matrilineal descent) as observed by Richards (1950) for African societies. In a phylogenetic comparative analysis of the coevolution of descent systems and cattle-keeping, Holden & Mace (2003) found evidence that Bantu matriliny was only sustained under certain socio-ecological conditions, i.e. the presence of horticulture and the absence of pastoralist subsistence systems. In this framework, both the prevalence of virilocality in ethnographically attested IE societies and the near-zero rate of switching from viri- to uxorilocality inferred by our evolutionary model are consistent with the pastoral and intensive agricultural subsistence economies ascribed to early IE societies (Mallory 1997). The matricentric character of AN societies (Burton *et al.* 1996) suggests a different evolutionary dynamic, that is, the loss of early—but perhaps widespread—uxorilocality. The origin and/or maintenance of uxorilocality has been linked to a 'male absence' factor (Keegan & Machlaclan 1989; Hage 1999). Many features of AN societies suggest this as a plausible hypothesis, including the unpredictable ecological features of oceanic environments; the voyaging traditions of seafaring people (both exploratory and trading-related); and subsistence systems that include deep-sea fishing but not pastoralism as practised on large continental landmasses. If variation in residence is indeed linked to a society's subsistence pattern and ecological niche, the type of analyses we present here offer good support for, and avenues for testing, the suspicions long held by anthropologists that human social life is not infinitely varied but rather is constrained by local environments. Asking the same questions in different ethnographic regions heralds a useful step forward in our ability to infer the general mechanisms of cultural evolutionary change, that is, the identification of lineage-specific processes within global domains (cf. Evans & Levinson 2009).

Investigating the evolution of cross-cultural diversity—in kinship or otherwise—involves an explicit choice about how to statistically approach hierarchically related human populations (Mace & Pagel 1994). As well as controlling for the effects of historical relatedness, phylogenetic comparative methods let us drill down into the specifics of ancestral states and processes of cultural change in a way that no other statistical methods currently available will allow. This is not, however, a statement that all human cultural processes follow strict 'vertical' or phylogenetic transmission dynamics (*contra* Borgerhoff Mulder *et al.* 2006). Rather we suggest, following Mace (2005), that questions about the degree to which cultural traits are transmitted 'vertically' from parent to daughter populations or 'horizontally' across populations only make sense within a phylogenetic framework. In this context, phylogenetic comparative methods have been shown to outperform non-phylogenetic methods under realistic scenarios and levels of horizontal transmission (Nunn *et al.* 2006; Currie *et al.* 2010). Ultimately, as with any other methodological approach, as long as the assumptions of the comparative analysis are made clear, the conclusions can be sustained or refuted by different data or analytical approaches.

Because the reconstruction of AN and IE prehistory are active fields of interdisciplinary scholarship, our findings have important implications for the interpretation of current ethnographic, archaeological, genetic and linguistic data; for example, fashionable statements in molecular anthropology about the impact of social structure on genetic diversity are largely used as post hoc narratives to explain incongruous findings. We believe there is obvious global utility in the methods and approaches presented here, and hope future research is stimulated by the promise of reconstructing the social lives of our ancestors.

REFERENCES

Anthony, D. W. 2007 *The horse, the wheel, and language: how Bronze-Age riders from the Eurasian steppes shaped the modern world.* Princeton, NJ: Princeton University Press.

Bentley, R. A., Price, T. D., Lüning, J., Gronenborn, D., Wahl, J. & Fullagar, P. D. 2002 Prehistoric migration in Europe: strontium isotope analysis of Early Neolithic skeletons. *Curr. Anthropol.* **43**, 799–804. (doi:10.1086/344373)

Blust, R. 1980 Early Austronesian social organization—the evidence of language. *Curr. Anthropol.* **21**, 205–247. (doi:10.1086/202430)

Boyd, R., Richerson, P. J., Borgerhoff-Mulder, M. & Durham, W. H. 1997 Are cultural phylogenies possible? In *Human by nature: between biology and the social sciences* (eds P. Weingart, P. J. Richerson, S. D. Mitchell & S. Maasen), pp. 355–386. Mahwah, NJ: Lawrence Erlbaum Associates.

Borgerhoff Mulder, M., Nunn, C. L. & Towner, M. C. 2006 Cultural macroevolution and the transmission of traits. *Evol. Anthropol.* **15**, 52–64. (doi:10.1002/evan.20088)

Burton, M. L., Moore, C. C., Whiting, J. W. M. & Romney, A. K. 1996 Regions based on social structure. *Curr. Anthropol.* **37**, 87–123. (doi:10.1086/204474)

Clackson, J. 2007 *Indo-European linguistics: an introduction.* Cambridge, UK: CUP.

Currie, T. E, Greenhill, S. J. & Mace, R. 2010 Is horizontal transmission really a problem for phylogenetic comparative methods? A simulation study using continuous cultural traits. *Phil. Trans. R. Soc. B* **365**, 3903–3912. (doi:10.098/rstb.2010.0014)

Diamond, J. & Bellwood, P. 2003 Farmers and their languages: the first expansions. *Science* **300**, 597–603. (doi:10.1126/science.1078208)

Dyen, I., Kruskal, J. B. & Black, P. 1992 An Indo-European classification: a lexicostatistical experiment. *Trans. Am. Phil. Soc.* **82**, 1–132.

Evans, N. & Levinson, S. C. 2009 The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–492. (doi:10.1017/S0140525X0999094X)

Felsenstein, J. 1985 Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15. (doi:10.1086/284325)

Fortunato, L. In press. Reconstructing the history of residence strategies in Indo-European-speaking societies: neo-, uxori-, and virilocality. *Hum. Biol.*

Fortunato, L. 2008 A phylogenetic approach to the history of cultural practices. In *Early human kinship: from sex to social reproduction* (eds N. J. Allen, H. Callan, R. Dunbar & W. James), pp. 189–199. Malden, MA: Blackwell Publishing Ltd.

Fortunato, L. & Mace, R. 2009 Testing functional hypotheses about cross-cultural variation: a maximum-likelihood comparative analysis of Indo-European marriage practices. In *Pattern and process in cultural evolution* (ed. S. Shennan), pp. 235–249. Berkeley, CA: University of California Press.

Fortunato, L., Holden, C. & Mace, R. 2006 From bridewealth to dowry? A Bayesian estimation of ancestral states of marriage transfers in Indo-European groups. *Hum. Nat.* **17**, 355–376. (doi:10.1007/s12110-006-1000-4)

Gimbutas, M. 1991 *The civilization of the goddess.* San Francisco, CA: HarperSanFrancisco.

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

Gray, R. D., Drummond, A. J. & Greenhill, S. J. 2009 Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323**, 479–483. (doi:10.1126/science.1166858)

Greenhill, S. J., Blust, R. & Gray, R. D. 2008 The Austronesian basic vocabulary database: from bioinformatics to lexomics. *Evol. Bioinform.* **4**, 271–283.

Guglielmino, C. R., Viganotti, C., Hewlett, B. & Cavalli-Sforza, L. L. 1995 Cultural variation in Africa: role of mechanisms of transmission and adaptation. *Proc. Natl Acad. Sci. USA* **92**, 7585–7589. (doi:10.1073/pnas.92.16.7585)

Haak, W. *et al.* 2008 Ancient DNA, strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. *Proc. Natl Acad. Sci. USA* **105**, 18 226–18 231. (doi:10.1073/pnas.0807592105)

Hage, P. 1998 Was Proto-Oceanic society matrilineal? *J. Polynesian Soc.* **107**, 365–379.

Hage, P. 1999 Reconstructing ancestral Oceanic society. *Asian Perspect.* **38**, 200–227.

Hage, P. & Marck, J. 2003 Matrilineality and the Melanesian origin of Polynesian Y chromosomes. *Curr. Anthropol.* **44**, S121–S127. (doi:10.1086/379272)

Harvey, P. H. & Pagel, M. D. 1991 *The comparative method in evolutionary biology.* Oxford, UK: Oxford University Press.

Holden, C. & Mace, R. 1997 Phylogenetic analysis of the evolution of lactose digestion in adults. *Hum. Biol.* **69**, 605–628.

Holden, C. J. & Mace, R. 2003 Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis. *Proc. R. Soc. Lond. B* **270**, 2425–2433. (doi:10.1098/rspb.2003.2535)

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. (doi:10.1126/science.1065889)

Hurles, M. E., Matisoo-Smith, E., Gray, R. D. & Penny, D. 2003 Untangling Oceanic settlement: the edge of the knowable. *Trends Ecol. Evol.* **18**, 531–540. (doi:10.1016/S0169-5347(03)00245-3)

Jordan, F. M., Gray, R. D., Greenhill, S. J. & Mace, R. 2009 Matrilocal residence is ancestral in Austronesian societies. *Proc. R. Soc. B* **276**, 1957–1964. (doi:10.1098/rspb.2009.0088)

Kayser, M., Lao, O., Saar, K., Brauer, S., Wang, X., Nürnberg, P., Trent, R. J. & Stoneking, M. 2008 Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *Am. J. Hum. Genet.* **82**, 194–198. (doi:10.1016/j.ajhg.2007.09.010)

Keegan, W. F. & Machlaclan, M. D. 1989 The evolution of avunculocal chiefdoms: a reconstruction of Taino kinship and politics. *Am. Anthropol.* **91**, 613–630. (doi:10.1525/aa.1989.91.3.02a00050)

Lane, R. B. 1961 A reconsideration of Malayo-Polynesian social organization. *Am. Anthropol.* **63**, 711–720. (doi:10.1525/aa.1961.63.4.02a00030)

LeBar, F. M. (ed.) 1975 *Ethnic groups of insular southeast Asia.* New Haven, CT: HRAF Press.

Levinson, D. (ed.) 1990 *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. (doi:10.1016/S0169-5347(00)02025-5)

Mace, R. 2005 On the use of phylogenetic comparative methods to test co-evolutionary hypotheses across cultures. In *The evolution of cultural diversity: a phylogenetic approach* (eds R. Mace, C. J. Holden & S. Shennan), pp. 235–256. London, UK: UCL Press.

Mace, R. & Pagel, M. 1994 The comparative method in anthropology. *Curr. Anthropol.* **35**, 549–564. (doi:10.1086/204317)

Mallory, J. P. 1997 Residence. In *Encyclopedia of Indo-European culture* (eds J. P. Mallory & D. Q. Adams), pp. 483–484. London, UK: Fitzroy Dearborn.

Marck, J. 2008 Proto Oceanic society was matrilineal. *J. Polynesian Soc.* **117**, 345–382.

Murdock, G. P. 1949 *Social structure.* New York, NY: MacMillan.

Murdock, G. P. 1967 *Ethnographic Atlas.* Pittsburgh, PA: University of Pittsburgh Press.

Nerlove, S. & Romney, S. K. 1967 Sibling terminology and cross-sex behavior. *Am. Anthropol.* **69**, 179–187. (doi:10.1525/aa.1967.69.2.02a00050)

Nunn, C. L., Borgerhoff Mulder, M. & Langley, S. 2006 Comparative methods for studying cultural trait evolution: a simulation study. *Cross-Cult. Res.* **40**, 177–209. (doi:10.1177/1069397105283401)

Pagel, M. 1994 Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. B* **255**, 37–45. (doi:10.1098/rspb.1994.0006)

Pagel, M. 1997 Inferring evolutionary processes from phylogenies. *Zool. Scripta* **26**, 331–348. (doi:10.1111/j.1463-6409.1997.tb00423.x)

Pagel, M. 1999 The maximum likelihood approach to reconstructing ancestral character states on phylogenies. *Syst. Biol.* **48**, 612–622.

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

Pagel, M. & Meade, A. 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* (eds R. Mace, C. J. Holden & S. Shennan), pp. 235–256. London, UK: UCL Press.

Pagel, M. & Meade, A. 2006 Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* **167**, 808–825. (doi:10.1086/503444)

Pagel, M., Meade, A. & Barker, D. 2004 Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53**, 673–684. (doi:10.1080/10635150490522232)

Pagel, M., Atkinson, Q. D. & Meade, A. 2007 Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* **449**, 717–720. (doi:10.1038/nature06176)

Price, T. D., Bentley, R. A., Lüning, J., Gronenborn, D. & Wahl, J. 2001 Prehistoric human migration in the Linearbandkeramik of Central Europe. *Antiquity* **75**, 593–603.

Raftery, A. E. 1996 Hypothesis testing and model selection. In *Markov chain Monte Carlo in practice* (eds W. R. Gilks, S. Richardson & D. J. Spiegelhalter), pp. 163–187. London, UK: Chapman & Hall/CRC.

Richards, A. R. 1950 Some types of family structure amongst the central Bantu. In *African systems of kinship and marriage* (eds A. R. Radcliffe-Brown & C. D. Forde), pp. 207–251. London, UK: Oxford University Press.

Ruhlen, M. 1991 *A guide to the world's languages: classification.* London, UK: Edward Arnold.

Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. 1998 Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **20**, 278–280. (doi:10.1038/3088)

Van Wouden, F. A. E. 1935 [1968] *Types of social structure in eastern Indonesia.* The Hague, The Netherlands: Nijhoff.

Wilkins, J. F. & Marlowe, F. 2006 Sex-biased migration in humans: what should we expect from genetic data? *Bioessays* **28**, 290–300. (doi:10.1002/bies.20378)