

The shape and tempo of language evolution

S. J. Greenhill^{1,2,*}, Q. D. Atkinson³, A. Meade⁴ and R. D. Gray¹

¹Department of Psychology, and ²Computational Evolution Group, Department of Computer Science, University of Auckland, Private Bag 92019, Auckland 1042, New Zealand

³Institute of Cognitive and Evolutionary Anthropology, University of Oxford, Oxford OX2 6PN, UK

⁴School of Biological Sciences, University of Reading, Reading RG6 6A7, UK

There are approximately 7000 languages spoken in the world today. This diversity reflects the legacy of thousands of years of cultural evolution. How far back we can trace this history depends largely on the rate at which the different components of language evolve. Rates of lexical evolution are widely thought to impose an upper limit of 6000–10 000 years on reliably identifying language relationships. In contrast, it has been argued that certain structural elements of language are much more stable. Just as biologists use highly conserved genes to uncover the deepest branches in the tree of life, highly stable linguistic features hold the promise of identifying deep relationships between the world's languages. Here, we present the first global network of languages based on this typological information. We evaluate the relative evolutionary rates of both typological and lexical features in the Austronesian and Indo-European language families. The first indications are that typological features evolve at similar rates to basic vocabulary but their evolution is substantially less tree-like. Our results suggest that, while rates of vocabulary change are correlated between the two language families, the rates of evolution of typological features and structural subtypes show no consistent relationship across families.

Keywords: language structure; language evolution; typology; evolution; phylogenetics; networks

1. INTRODUCTION

How far back can we trace the history of languages? The traditional comparative method in historical linguistics uses systematic sound correspondences between homologous ('cognate') words to infer relatedness between languages. Most linguists argue that this approach can only be used to make inferences about languages that diversified within the last 6000–10 000 years (Nichols 1992; Ringe 1995; Kaufman & Golla 2000). Beyond this time, however, it becomes impossible to distinguish accurately whether any signal in the data represents descent from a common ancestor or false similarities owing to chance and borrowing between languages.

Some authors have claimed that certain typological features that describe the structures present in a language, such as ergativity, head marking and numeral classifiers, are more stable than the lexicon (Nichols 1992, 1994). If some typological features are consistently stable within language families, and resistant to borrowing, then they might hold the key to uncovering relationships at far deeper levels than previously possible. For example, Nichols (1994) uses typological features to argue for a spread of languages and cultures around the Pacific Rim, connecting Australia, Papua New Guinea, Asia, Russia, Siberia, Alaska and the western coasts of North and South America. If this is correct, then these typological features must be reflecting time depths of at least 16 000 years and possibly as deep as 50 000 years ago (Nichols 1994). A recent phylogenetic study of phonological and morphosyntactic features in non-Austronesian

languages of Island Melanesia argued that typological traits reveal a phylogenetic signal consistent with deep (approx. 10 000 years) historical relationships (Dunn *et al.* 2005). One explanation for this stability is that the evolution of typological features is more constrained than that of the lexicon because structural traits function as an interrelated system with strong dependencies between components ('un système où tout se tient', variously attributed to Antoine Meillet, and Ferdinand de Saussure; Peeters 1990).

However, the lack of comprehensive worldwide typological data has made it difficult to assess the overall shape and tempo of changes in language structure. The recently published *World atlas of language structures* (WALS) remedies this problem (Haspelmath *et al.* 2005). WALS includes information about 141 typological features from 2561 languages. Here, we report the results of phylogenetic analyses of the typological data in the WALS. First, we explore the global pattern of typological data using a network method to assess evidence for a deep signal in the data. Second, we quantify the fit of typological and lexical features onto known family trees for two of the world's largest and best-studied language families—Indo-European and Austronesian. Third, we infer the rates of evolution of typological and lexical features within these families and compare rates between families.

2. MATERIAL AND METHODS

(a) *Typological data*

From the 141 characters in the WALS (Haspelmath *et al.* 2005), we discarded the three characters belonging to the 'sign languages' and 'other' categories, leaving 138 characters for analysis (electronic supplementary material, table S1). We extracted three datasets from WALS. The first dataset

* Author for correspondence (s.greenhill@auckland.ac.nz).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2010.0051> or via <http://rspb.royalsocietypublishing.org>.

was a ‘worldwide’ dataset that included all languages in WALS with less than 25 per cent missing data (electronic supplementary material, table S2). Unfortunately, the WALS database has incomplete data for many languages and feature classes, so this left a total of 99 languages in this worldwide dataset. The second and third datasets comprised 20 Austronesian and 20 Indo-European languages that we had sufficient lexical data for and that were well described in the WALS database. To maximize the phylogenetic signal in the typological data, we recoded 49 of the 138 characters by splitting up aggregate categories and combining feature states with few members (see electronic supplementary material and table S3).

(b) *Lexical data*

Lexical cognate data for the languages in WALS were taken from two sources (electronic supplementary material, tables S4 and S5). The Austronesian lexical data were extracted from the Austronesian Basic Vocabulary Database (Greenhill *et al.* 2008; <http://language.psy.auckland.ac.nz/austronesian>). This database project contains 210-item wordlists and cognate information from over 650 Austronesian languages. The Indo-European lexical data came from a published dataset of 200-item basic vocabulary wordlists and cognate information from 95 Indo-European languages (Dyen *et al.* 1992). Both the Austronesian and Indo-European databases comprised items of basic vocabulary (terms for body parts, kinship terms, colours, simple verbs, numbers, etc.) that are thought to be highly stable over time and resistant to being borrowed between languages (Swadesh 1952).

(c) *NeighbourNet analysis*

The worldwide NeighbourNet was constructed using SPLITS TREE v4.8 using uncorrected P-distances (Bryant & Moulton 2004; Bryant *et al.* 2005). To reduce the noise in the network, splits were filtered according to a weight threshold of 0.002. NeighbourNets were also constructed for each typological/lexical dataset using the same method, and splits were filtered to a threshold of 0.001 (electronic supplementary material, figure S5).

(d) *Character fit analysis*

We constructed family trees for the Indo-European and Austronesian language families (electronic supplementary material, figure S4) from the standard *Ethnologue* classification (Gordon 2005) and previous research on Indo-European (Gray & Atkinson 2003; Atkinson & Gray 2005), and Austronesian (Blust 1999; Lynch *et al.* 2002; Gray *et al.* 2009). To measure the fit of each character onto these trees, we calculated the retention index (RI; Archie 1989; Farris 1989) for all characters in the four datasets (Austronesian Lexicon, Austronesian Typology, Indo-European Lexicon and Indo-European Typology) using PAUP* v.4b10 (Swofford 2002). We selected the RI for this comparison as it does not require us to estimate branch lengths as likelihood-based character-fit analyses would. RIs are only available for characters that are parsimony informative (constant characters or characters with all unique states do not provide information on the fit of the data to a tree). RIs were calculated for 113/210 characters in the Austronesian lexicon, 109/138 characters in the Austronesian typology, 183/200 characters in the Indo-European lexicon, and 116/138 characters in the Indo-European typology.

(e) *Rates analysis*

To calculate the rate estimates, trees with branch lengths proportional to the amount of change between each language are required. We used a Bayesian phylogenetic approach implemented in the program BAYESPHYLOGENIES (Pagel & Meade 2004) to produce a posterior distribution of phylogenetic trees from the binary-coded lexical cognate data. The analysis used a two-rate model of cognate evolution that allows cognates to be gained and lost at different rates. The Markov chain ran for 10 million generations, and burn-in was set to 5 million generations after inspection of log likelihood plots of the parameters. The tree topologies were constrained to match the classification trees (electronic supplementary material, figure S4), so that each tree sample varied only in their estimate of the branch lengths. Trees were sampled every 5000 generations from the chain, leaving a total of 1000 post-burn-in trees.

By constraining the tree topology to established language groupings we minimize any bias that might result from estimating the tree topology from the lexical cognate data. The use of the lexical data to estimate the branch lengths is consistent with arguments that lexical phylogenies based on basic vocabulary provide good estimators of the underlying cultural history (Mace & Pagel 1994). Moreover, the site-specific likelihoods (indicating the fit of the data under the model of evolution) calculated on the trees with branch lengths derived from the typological data were essentially identical to those obtained with lexical branch lengths (Spearman's $\rho = 0.997$, $p < 0.001$)—in other words, there is no reason to think that the use of lexical branch lengths biases our results.

Maximum-likelihood rate estimates, μ , were calculated from these posterior tree distributions using BAYESTRAITS (Pagel *et al.* 2004). BAYESTRAITS implements a continuous time Markov model that allows characters to change between states over small time intervals. This can be used to reconstruct how traits with discrete, finite states evolve on the trees in the posterior distribution. Estimates of μ were obtained for all four datasets (Austronesian lexicon, Austronesian typology, Indo-European lexicon and Indo-European typology). Traits with greater than 50 per cent missing data were excluded from the analyses. For constant characters, the maximum-likelihood rate estimate is zero. However, for any trait that can vary, the true rate is always non-zero. We can infer a rate for constant characters by plotting the observed number of states against the rate estimates for each feature within each of the four datasets. We fitted an exponential curve to the data and used this to provide a predicted rate for constant characters in each dataset—the point on the curve where the observed number of states is one. The results we report include the estimated rates for non-constant characters and the inferred rate for constant characters. We also repeated all rate analyses setting the constant rate to the minimum estimated maximum-likelihood rate among the variable characters. This had no appreciable effect on the results we report.

3. RESULTS

(a) *The global pattern of typological diversity*

To explore global patterns of the typological signal, we used a phylogenetic network technique, NeighbourNet (Bryant & Moulton 2004; Bryant *et al.* 2005), to visualize the relationships implied by these data (figure 1). In these

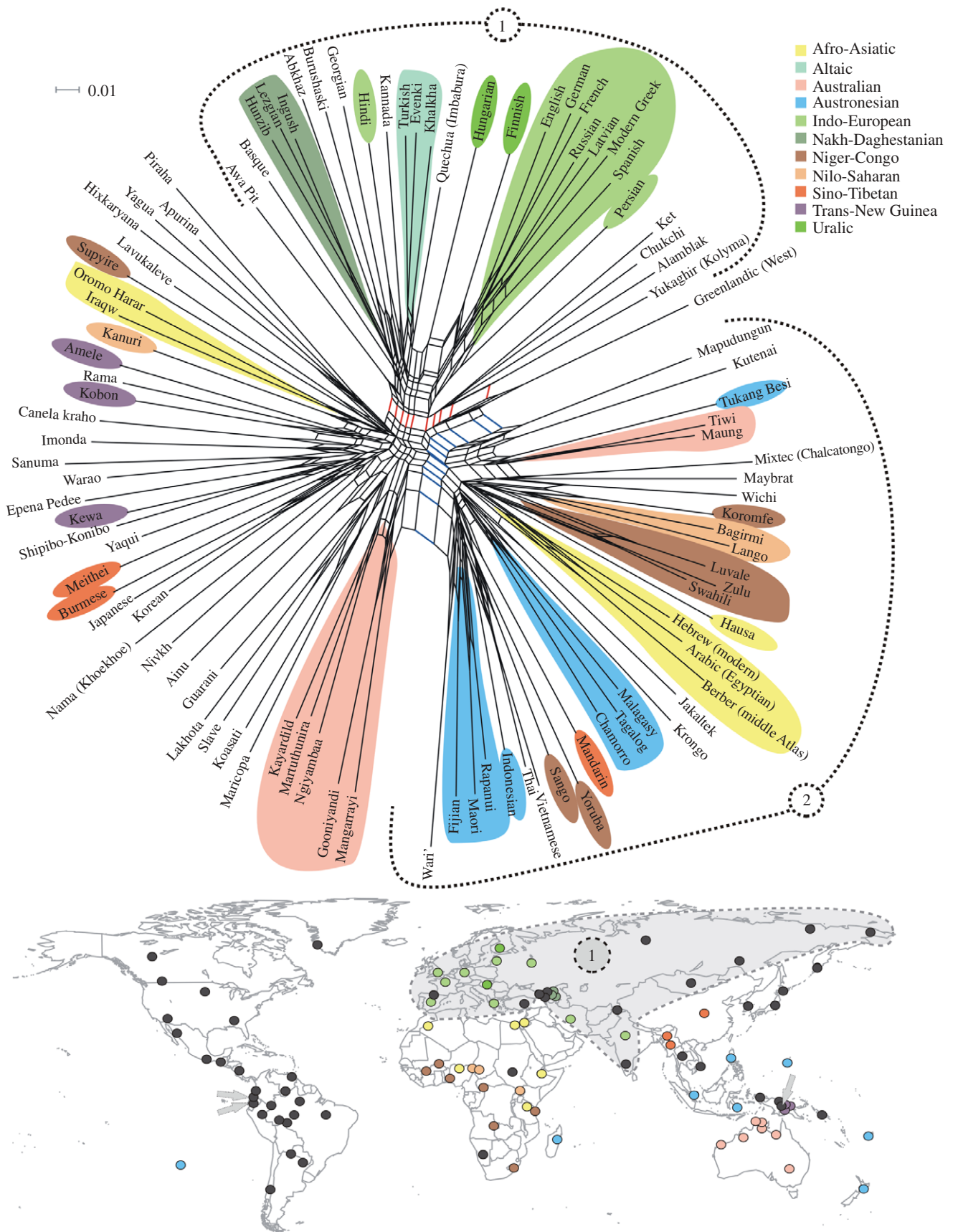


Figure 1. NeighbourNet for the 99 most well-attested languages in the WALS database. This network is based on 138 typological characters and shows the signals grouping languages. Branch lengths are proportional to amount of divergence between languages, and the box-like structures reflect conflicting signal. Accepted family groups are colour-coded, and potential language areas are marked with dashed lines and numbered as described in the text. The dashed area and arrows on the map show the extent of the large Eurasian cluster (1). Cluster 2 appears to be a residual grouping containing languages from Australia, Africa and the Pacific.

networks, the length of the branches is proportional to the amount of divergence between languages. Box-like structures represent the conflicting signals when typological features support incompatible language groupings. If typological features are deeply stable, then we would expect the groupings in the network to reflect known linguistic history and contain few boxes of conflicting signals. In contrast, if the typological features tend to diffuse between adjacent languages in a linguistic area or evolve too rapidly to reveal a deep signal, we would expect to see a star-like network with many boxes and clusters reflecting geographical proximity or chance resemblances.

The network in figure 1 correctly groups some of the languages into known language families, with Indo-European, Altaic and Nakh-Daghestanian being the most distinct. The network also groups a number of subfamilies together—such as the Pama-Nyungan languages (Kayardild, Martuthunira and Ngiyambaa), the Bantu languages (Luvale, Swahili and Zulu), the Oceanic languages (Maori, Fijian and Rapanui), the Semitic languages (Hebrew and Arabic) and the Cushitic languages (Irakw and Oromo Harar). However, other well-known families are not recovered, including Sino-Tibetan, Uralic, and Trans-New Guinea. The Austronesian language family also does not form a monophyletic group. Additionally, the network shows evidence of a substantial conflicting signal between structural elements (box-like structures) and does not accurately recover many attested phylogenetic relationships within the major language families. For example, in Indo-European, the network links German to French, when German is more closely related to English (Beekes 1995).

The network does, however, show evidence for some higher level clusters in the data. The first of these (cluster 1, labelled in figure 1) includes the languages from continental Eurasia, which could be interpreted as indicating an ancient common ancestry. This cluster groups the Indo-European languages with the Uralic languages (Finnish and Hungarian), consistent with the proposed macro-family Indo-Uralic. These two families are joined in this cluster by the Altaic language family (Turkish, Evenki and Khalkha), the Dravidian language Kannada and a number of languages from the Caucasus region: the Nakh-Daghestanian family (Ingush, Lezgian and Hunzib), Abkhaz (Northwest Caucasian) and Georgian (Kartvelian). If typological features do indeed evolve slowly enough to reveal a deep history, then this cluster may represent the controversial Nostratic macro-family (Renfrew & Nettle 1999). However, the inclusion of languages such as Alamlak (from Papua New Guinea), Awa Pit (from Colombia), Quechua (from Ecuador) and the isolate Basque are incompatible with this proposal. A second large cluster (cluster 2, labelled in figure 1) includes the Australian languages, the Austronesian languages, and some languages from the African families of Afro-Asiatic and Niger-Congo. This second cluster does not correspond to any known macro-family proposals or geographical regions, however, Austronesian languages are placed next to some other non-Austronesian languages from Southeast Asia (Thai, Vietnamese and Mandarin).

The left side of the network (figure 1) contains a subset of the languages of Australia, and distinguishes between

the Pama-Nyungan languages (Kayardild, Martuthunira, Ngiyambaa), and others from different families (Gooniyandi, Mangarayi). However, two other languages from the northern tip of Australia (Tiwi and Maung) are not included but placed in the second cluster. Another interesting subset here may also hint at some deeper links—most of the languages of North America are linked together in this network (Lakhota, Slave, Maricopa and Koasati). However, this grouping rather unusually includes a language from Paraguay—Guarani—and does not include other North American languages of Yaqui and Kutenai.

(b) *Modelling structural and lexical evolution on trees*

The existence of high-level clusters in the WALS data is consistent with the proposal that some typological features evolve slowly enough to identify deep historical relationships. However, phylogenetic networks cannot distinguish between similarity owing to common ancestry and similarity owing to areal diffusion or chance resemblances arising through independent innovation. To evaluate the claim that some typological features of language are highly stable, we compared the shape and tempo of typological and lexical evolution by modelling their replacement through time on two language family trees that have well-established internal subgroupings: Indo-European (Beekes 1995; Gray & Atkinson 2003; Atkinson & Gray 2006), and Austronesian (Blust 1999; Lynch *et al.* 2002; Gray *et al.* 2009). If some typological features are highly stable and good indicators of common ancestry, then we would expect them (i) to fit well with established language groupings and (ii) to show slower rates of change than lexical features as a whole. We extracted typological data from the WALS for the 20 most well-attested languages in each of the two families, removing the languages with the least data. We assembled lexical datasets for the same 20 languages from published databases of the Indo-European (Dyen *et al.* 1992) and Austronesian (Greenhill *et al.* 2008) vocabulary.

We assessed the shape of language evolution in these data by estimating the fit of the typological and lexical data onto the established family trees using the RI (Archie 1989; Farris 1989). A stable, well-fitting character will have an RI approaching one, while an unstable or rapidly evolving character will have an RI approaching zero. Histograms of the RIs for the lexical and typological features in the Indo-European and Austronesian datasets are shown in figure 2*a*. In the lexical data, the mean RI for each character was 0.84 (s.d. = 0.31) for the Austronesian and 0.89 (s.d. = 0.21) for the Indo-European vocabulary. The mean RI per character of the typological data was much lower at 0.36 (s.d. = 0.33) for the Austronesian and 0.32 (s.d. = 0.33) for the Indo-European. In both families, the lexical data were a significantly better fit to the expected family trees than the typological data (Mann-Whitney: Austronesian $U = 8331$, $p < 0.001$, Indo-European $U = 13\,086.5$, $p < 0.001$). These differences in fit are also evident in networks of the typological and lexical data (figure 2*a*, inset) where the lexical networks clearly show a much more tree-like signal than the typological networks. Unfortunately, the RI is unable to estimate the fit of constant characters on the trees.

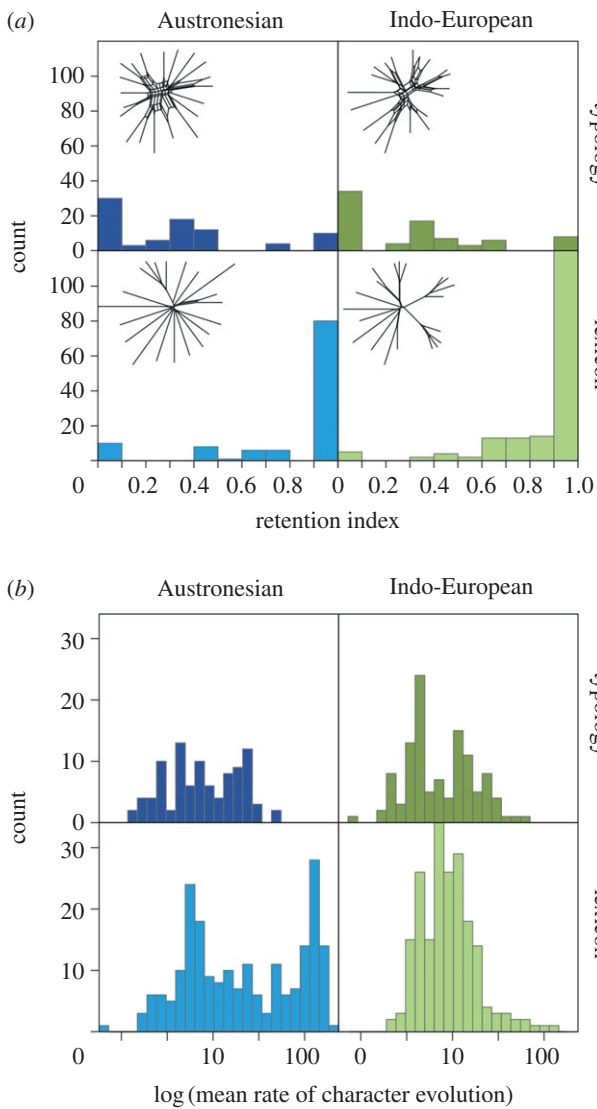


Figure 2. Histograms comparing the Austronesian and Indo-European lexical and typological data. (a) The retention index (RI) for each character state in the typological and lexical characters on the established classification trees. NeighbourNets of each dataset are inset in (a), showing the greater amounts of conflicting signal found in both typological datasets. (b) The frequency distribution of observed rates of evolution across all the characters in both families and data types.

The characters that are constant in both language families ($n = 6$) are potential candidates for deep relationship indicators. However, a closer inspection of these characters shows that four of them are only constant owing to large numbers of missing data (with only approx. 12.5% of the states assigned across the 40 languages). The two characters that are constant in both families and have appreciable amounts of data are ‘N-M Pronouns’ (with 15 of the 40 languages showing ‘no N-M pronouns’, and the remainder missing data) and ‘order of adverbial subordinator and clause’ (with 38 of 40 languages belonging to the state ‘adverbial subordinators which are separate words and which appear at the beginning of the subordinate clause’).

It could be argued that the analysis of character fit is biased in favour of the lexical cognate data since historical linguistics often uses lexical information to infer linguistic

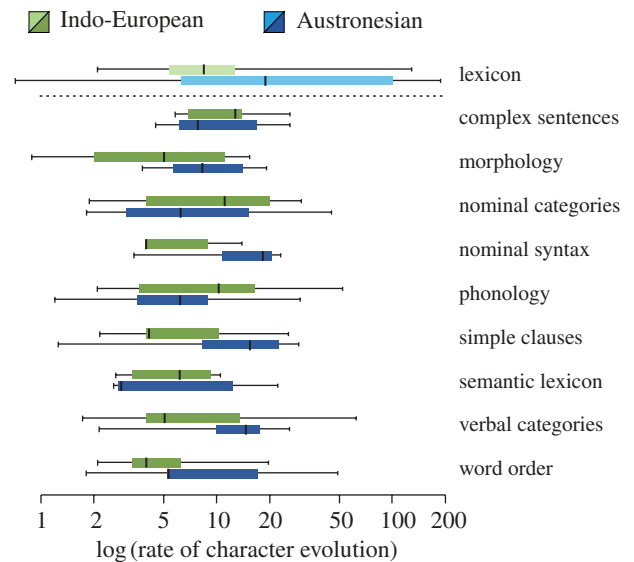


Figure 3. Boxplots showing the observed rate of change by feature class across the Austronesian/Indo-European language families. Values closer to zero are evolving more slowly.

relationships. Indeed, some subgroups are defined by major lexical innovation, such as Eastern Malayo-Polynesian (Blust 2009). In other cases, however, subgroups are defined by phonological and morphological innovations (Durie & Ross 1996; Blust 2009). For example, the Proto-Nuclear Polynesian subgroup is demarcated by many morphological innovations, the Oceanic subfamily is defined by the phonological merger of *p and *b, and Central-Eastern Malayo-Polynesian is identified by the lowering of high vowels and four shared grammatical morphemes (Blust 2009). The subgroups we use here represent the best available estimate of the true underlying language tree, drawing on a consilience of evidence from both lexical and structural data (Durie & Ross 1996; Blust 2009). Any bias in favour of the cognate data is therefore expected to be minimal.

To estimate rates of change, we calculated the maximum-likelihood estimate for the rate of evolution across the posterior distribution of trees in each family. Figure 3 shows a comparison of the distributions of rates for Indo-European and Austronesian lexical and typological characters. In both families, the distributions of lexical and typological rates are comparable. The similar ranges evident in these plots indicate that there is in fact no substantial difference between the slowest rates of lexical and typological change in either family. Austronesian rates for lexical features were on average slightly higher than rates for typological features (Mann–Whitney: Austronesian $U = 5961$, $p < 0.001$) while in the Indo-European data, lexical and typological rates were not significantly different (Mann–Whitney: Indo-European $U = 6718$, $p > 0.05$). The bimodal distribution for Austronesian lexicon indicates that its higher average rate is due to a relatively high number of rapidly evolving words.

While we find no clear difference between overall rates of lexical and typological change, some subsets of typological features may nonetheless change slowly enough to

infer deep relationships. For example, Nichols (1992) claims that ergativity, head marking and numeral classifiers are among the most stable structural features of language. The WALS project groups the typological data into nine feature classes describing different aspects of language structure. Figure 3 shows the inferred rate distributions grouped according to the nine typological feature classes defined in the WALS database, together with the lexical rate estimates. This plot highlights considerable variation in the rates of evolution between feature classes and between families. For example, characters in the nominal syntax feature class have some of the highest rates in Austronesian but lowest in Indo-European, while the reverse is true for complex sentence structures. A univariate ANOVA shows that, when controlling for language family, there is no effect of typological feature class on rates of feature evolution ($F = 1.27$, $p = 0.26$).

Finally, we examined the relationship between rates of change for individual lexical and typological features across language families. Identifying specific features that are consistently stable across families has the potential to greatly improve our ability to detect and evaluate deep inter-family relationships. In addition, the kinds of regularities identified may point to constraints on the process of language evolution itself. In agreement with previous research (Pagel 2000; Pagel & Meade 2006), we find that rates of lexical change are correlated across language families (Spearman's $\rho = 0.37$, $p < 0.001$). By contrast, there is no significant correlation in rates of typological feature change between Indo-European and Austronesian (Spearman's $\rho = 0.17$, $p = 0.10$). Although non-significant, this relationship is positive, suggesting a small number of structural features may still be consistently stable. We can identify nine features that have rates in the slowest 0.20 quantile in both language families: the velar nasal, case syncretism, numeral bases, pronominal and adnominal demonstratives, the optative, coding of nominal plurality, glottalized consonants, syllable structure and suppletion according to tense and aspect. These traits could be seen as candidates for investigating deep time scales; however, caution is needed in interpreting these results. First, a χ^2 -test reveals that finding nine traits in the slowest 0.20 percentile in both families does not differ significantly from chance ($\chi^2 = 3.487$, $p = 0.062$), and the same applies using the 0.05 percentile ($\chi^2 = 2.34$, $p = 0.13$). Second, many of these characters reflect shared absence in the majority of the languages in our sample. For example, for the character the optative, WALS only has data for 30/40 of the languages in our sample, and 28 of these are marked as 'inflectional optative absent'. Likewise, in the character the velar nasal, the Austronesian languages show their well-known bias for nasal substitution (Blust 2004), with 11 of the 20 languages having initial velar nasals, eight languages missing data and only Kilivila showing an absence. However, in the 12 Indo-European languages with data, the most prominent state (10/12) is 'no velar nasal'. Together with the absence of any correlation in the typological rates of evolution between the families, these patterns do not support the existence of a set of universally stable typological features.

4. DISCUSSION

There is considerable interest in the possibility that analyses of typological features may enable us to 'push back the time barrier' beyond the apparent 6000–10 000 year upper limit of the comparative method (Gray 2005). It has been suggested that typology can reveal historical signal dating back at least this far (Dunn *et al.* 2005, 2008), or even tens of thousands of years earlier (Nichols 1994). The network analysis of WALS structural features reported in figure 1 points to some intriguing possible deep relationships, perhaps most notably the cluster linking together many of the major language families of Eurasia. However, our analysis of rates of evolution failed to identify any typological features that evolve at consistently slower rates than the basic lexicon. If the signal in the lexicon does stretch back as far as 10 000 years (Nichols 1992; Ringe 1995; Kaufman & Golla 2000), then our results suggest that typological data is constrained by a similar time horizon (e.g. Dunn *et al.* 2005, 2007, 2008).

Beyond the difficulty of identifying consistently stable typological features, our findings suggest two further challenges to inferring deep ancestral relationships from structural language data. First, the typological features show relatively high rates of homoplasy. The classification of lexical data into cognate sets relies on isomorphism between sound and meaning within a vast possible state space of the items under comparison. The coupling of these two aspects reduces the possibility of chance similarity (Meillet 1948). In contrast, there is a 'poverty of choice' of possible typological states (Harrison 2003). For example, there are only six permutations for the ordering of the subject, object and verb that a language can use. Accordingly, there is a 1/6 chance that any two languages share the same ordering—in fact, since some configurations are much more likely than others, even this probability is an underestimate. This means that, even for a given rate of change, shared typological features are a less reliable indication of a common ancestry than shared basic vocabulary, and are more likely to produce spurious relationships.

A second issue with identifying slowly evolving typological features is diffusion between geographically proximate languages (Matras *et al.* 2006). This can occur through processes like language shift (Thomason & Kaufman 1988)—where speakers of one language change to another owing to societal influences, yet retain morphology or phonology from their original language, or metatypy (Ross 1996)—where a language rearranges some aspect of typology (e.g. morphosyntax) owing to contact between languages without explicit borrowing between the languages, usually as an outcome of intimate cultural contact. Our results show a substantial non-tree-like signal in the typological data and a poor fit with known language relationships within the Austronesian and Indo-European language families. On a global scale, figure 1 shows some putative geographical clusters like the 'Nostratic' grouping in Eurasia. In this Nostratic cluster, Hindi does not group correctly with Indo-European but is located with its geographical neighbour, the Dravidian language, Kannada, suggesting that the similarities seen here may indeed be due to diffusion. Likewise, a grouping of Indonesian, Thai, Vietnamese and Mandarin may be the result of areal diffusion in the Southeast Asian region

(Bisang 2006; Matras *et al.* 2006). The areal diffusion of typological features—like lexical borrowing—does make it harder to identify common ancestry.

Diffusion and chance resemblances are serious challenges for historical inference based on typological data. The problem of diffusion can be lessened if known instances of diffusion are identified and removed (Ross 1996; Dunn *et al.* 2008), and the data are analysed with methods that are robust to the effect of diffusion (Greenhill *et al.* 2009). For example, the WALS contains information about word order (subject, object and verb), but additional distinctions can be made between word order for different kinds of clauses (e.g. main versus subordinate clauses) or between clausal and nominal objects. By identifying these and other more specific character states, it may be possible to increase the historical signal in typological data (Reesink *et al.* 2009), although rates of evolution will then necessarily increase. In addition, the WALS data is unfortunately sparse, containing only 138 characters (compared with the approx. 200 well-attested items of lexicon), and with many languages missing information—perhaps more signal will be evident in a more complete dataset.

While we were unable to identify a set of consistently stable typological features, rates of lexical evolution in one family were a good predictor of rates in the other. This fits with previous work showing that rates of change in lexical items are highly correlated across the Indo-European, Austronesian and Bantu language families (Pagel 2000; Pagel & Meade 2006). Recent work has also shown that rates of lexical change are predictable based on the frequency of use and part of speech (Pagel *et al.* 2007) and that some meanings have a lexical ‘half-life’—the time after which there is a 50 per cent chance that the word is replaced—in excess of 20 000 years. These extremely slow and predictable rates of lexical change mean that basic vocabulary may be a more practical choice for investigating questions of deeper language origins.

Finally, our findings highlight how little we know about the shape and tempo of language change. Contrary to what might be intuitively expected, our results indicate that dependencies between structural elements of language appear to do little to slow down rates of structural change, or to limit the diffusion of features between languages. In addition, we find that rates of structural evolution are specific to each language family, while lexical rates are correlated across families. One explanation for this observation may be that the frequency of use of different structural elements is an important determinant of rates of structural change, just as is the case for lexical change (Pagel *et al.* 2007). While frequency of word use is relatively constant across languages, the way structures are used depends on what other structural constraints operate in a language (Meillet 1948). This may explain the variation we see in rates of structural evolution between language families. In future, model-based approaches like those outlined here could be used to test hypotheses about macro-scale language change, and so shed light on the basic mechanisms driving the shape and tempo of language evolution.

Funding was provided by a Bright Futures Top Achiever Doctoral Scholarship to S.G. and by a Royal Society of New Zealand Marsden Grant to R.G. and S.G. We would

like to thank David Bryant, Lyle Campbell, Tom Currie, Michael Dunn, Neil Gemmell, Mark Pagel, Malcolm Ross, Robert Ross, Annik van Toledo and two anonymous reviewers for discussion. We thank the Centre for Advanced Computing and Emerging Technologies (ACET) at the University of Reading for making the ThamesBlue supercomputer available for our use.

Conceived and designed the experiments: S.G., Q.A., R.G. Performed the experiments: S.G., Q.A., A.M. Analysed the data: S.G., Q.A., A.M. Contributed analysis tools: A.M. Wrote the paper: S.G., Q.A., R.G.

REFERENCES

- Archie, J. W. 1989 Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. *Syst. Zool.* **38**, 253–269. (doi:10.2307/2992286)
- Atkinson, Q. D. & Gray, R. D. 2006 How old is the Indo-European language family? Progress or more moths to the flame? In *Phylogenetic methods and the prehistory of languages* (eds P. Forster & C. Renfrew), pp. 91–109. Cambridge, UK: McDonald Institute for Archaeological Research.
- Beekes, R. S. P. 1995 *Comparative Indo-European linguistics*. Amsterdam, The Netherlands: John Benjamins.
- Bisang, W. 2006 Contact-induced convergence: typology and areality. In *Encyclopedia of language and linguistics*, vol. 3 (ed. K. Brown), pp. 88–101. Oxford, UK: Elsevier.
- Blust, R. A. 1999 Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In *Selected Papers from the Eighth Int. Conf. on Austronesian Linguistics*, vol. 1 (eds E. Zeitoun & P. Li), pp. 31–94. Taipei, Taiwan: Academia Sinica.
- Blust, R. 2004 Austronesian nasal substitution: a survey. *Ocean. Linguist.* **43**, 73–148. (doi:10.1353/ol.2004.0004)
- Blust, R. 2009 *The Austronesian languages*. Canberra, Australia: Pacific Linguistics.
- Bryant, D. & Moulton, V. 2004 Neighbour-Net: an agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21**, 255–265. (doi:10.1093/molbev/msh018)
- Bryant, D., Filimon, F. & Gray, R. D. 2005 Untangling our past: languages, trees, splits and networks. In *The evolution of cultural diversity: phylogenetic approaches* (eds R. Mace, C. J. Holden & S. Shennan), pp. 67–84. London, UK: UCL Press.
- Dunn, M., Terrill, A., Reesink, G., Foley, R. A. & Levinson, S. C. 2005 Structural phylogenetics and the reconstruction of ancient language history. *Science* **309**, 2072–2075. (doi:10.1126/science.1114615)
- Dunn, M., Foley, R., Levinson, S., Reesink, G. & Terrill, A. 2007 Statistical reasoning in the evaluation of typological diversity in Island Melanesia. *Ocean. Linguist.* **46**, 388–403.
- Dunn, M., Levinson, S. C., Lindström, E., Reesink, G. & Terrill, A. 2008 Structural phylogeny in historical linguistics: methodological explorations applied in Island Melanesia. *Language* **84**, 710–759. (doi:10.1353/lan.0.0069)
- Durie, M. & Ross, M. 1996 *The comparative method reviewed: regularity and irregularity in language change*. New York, NY: Oxford University Press.
- Dyen, I., Kruskal, J. B. & Black, P. 1992 An Indo-European classification: a lexicostatistical experiment. *Trans. Am. Phil. Soc.* **82**, iii–132. (doi:10.2307/1006517)
- Farris, J. 1989 The retention index and the rescaled consistency index. *Cladistics* **5**, 417–419. (doi:10.1111/j.1096-0031.1989.tb00573.x)

- Gordon, R. G. J. (ed.) 2005 *Ethnologue: languages of the world*. Dallas, TX: SIL International.
- Gray, R. 2005 Pushing the time barrier in the quest for language roots. *Science* **309**, 2007–2008. (doi:10.1126/science.1119276)
- Gray, R. D. & Atkinson, Q. D. 2003 Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* **426**, 435–439. (doi:10.1038/nature02029)
- 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.
- Greenhill, S. J., Currie, T. E. & Gray, R. D. 2009 Does horizontal transmission invalidate cultural phylogenies? *Proc. R. Soc. B* **276**, 2299–2306. (doi:10.1098/rspb.2008.1944)
- Harrison, S. P. 2003 On the limits of the comparative method. In *The handbook of historical linguistics* (eds B. D. Joseph & R. D. Janda), pp. 213–243. Malden, MA: Blackwell.
- Haspelmath, M., Dryer, M., Gil, D. & Comrie, B. (eds) 2005 *The world atlas of language structures*. Oxford, UK: Oxford University Press.
- Kaufman, T. & Golla, V. 2000 Language groupings in the new world: their reliability and usability in cross-disciplinary studies. In *America past, America present: genes and languages in the Americas and beyond* (ed. C. Renfrew), pp. 47–57. Cambridge, UK: The McDonald Institute for Archaeological Research.
- Lynch, J., Ross, M. & Crowley, T. 2002 *The oceanic languages*, Curzon Language Family Series. Richmond, UK: Curzon.
- Mace, R. & Pagel, M. 1994 The comparative method in anthropology. *Curr. Anthropol.* **35**, 549–564. (doi:10.1086/204317)
- Matras, Y., McMahon, A. & Vincent, N. 2006 *Linguistic areas: convergence in historical and typological perspective*. New York, NY: Palgrave.
- Meillet, A. 1948 *Linguistique historique et linguistique et générale*. Paris, France: Champion.
- Nichols, J. 1992 *Linguistic diversity in space and time*. Chicago, IL: University of Chicago Press.
- Nichols, J. 1994 The spread of language around the Pacific Rim. *Evol. Anthropol.* **3**, 206–215. (doi:10.1002/evan.1360030607)
- Pagel, M. 2000 Maximum-likelihood models for glotto-chronology and for constructing linguistic phylogenies. In *Time depth in historical linguistics*, vol. 1 (eds C. Renfrew, A. McMahon & L. Trask), pp. 189–207. Cambridge, UK: McDonald Institute for Archaeological Research.
- 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. 2006 Estimating rates of lexical replacement on phylogenetic trees of languages. In *Phylogenetic methods and the prehistory of languages* (eds C. Renfrew & P. Forster), pp. 173–182. Cambridge, UK: McDonald Institute Monographs.
- 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)
- Peeters, B. 1990 Encore une fois ‘où tout se tient’. *Hist. Ling.* **17**, 427–463.
- Reesink, G., Singer, R. & Dunn, M. 2009 Explaining the linguistic diversity of Sahul using population models. *PLoS Biol.* **7**, e1000241. (doi:10.1371/journal.pbio.1000241)
- Renfrew, C. & Nettle, D. 1999 *Nostratic: examining a linguistic macrofamily*. Cambridge, UK: McDonald Institute for Archaeological Research.
- Ringe, D. 1995 ‘Nostratic’ and the factor of chance. *Diachronica* **12**, 55–74.
- Ross, M. D. 1996 Contact-induced change and the comparative method: cases from Papua New Guinea. In *The comparative method reviewed: regularity and irregularity in language change* (eds M. Durie & M. D. Ross), pp. 180–217. New York, NY: Oxford University Press.
- Swadesh, M. 1952 Lexico-statistic dating of prehistoric ethnic contacts. *Proc. Am. Phil. Soc.* **96**, 452–463.
- Swofford, D. L. 2002 *PAUP*: phylogenetic analysis using parsimony (* and other methods)*. Sunderland, MA: Sinauer Associates.
- Thomason, S. G. & Kaufman, T. 1988 *Language contact, creolization, and genetic linguistics*. Berkeley, CA: University of California Press.