A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone

Dan Dediu*

Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

Language is a hallmark of our species and understanding linguistic diversity is an area of major interest. Genetic factors influencing the cultural transmission of language provide a powerful and elegant explanation for aspects of the present day linguistic diversity and a window into the emergence and evolution of language. In particular, it has recently been proposed that *linguistic tone*—the usage of voice pitch to convey lexical and grammatical meaning—is biased by two genes involved in brain growth and development, *ASPM* and *Microcephalin*. This hypothesis predicts that tone is a stable characteristic of language because of its 'genetic anchoring'. The present paper tests this prediction using a Bayesian phylogenetic framework applied to a large set of linguistic features and language families, using multiple software implementations, data codings, stability estimations, linguistic classifications and outgroup choices. The results of these different methods and datasets show a large agreement, suggesting that this approach produces reliable estimates of the stability of linguistic data. Moreover, linguistic tone is found to be stable across methods and datasets, providing suggestive support for the hypothesis of genetic influences on its distribution.

Keywords: linguistic tone; genetic biasing; language phylogenies

1. INTRODUCTION

The approximately 7000 languages currently spoken around the world [1] vary enormously not only in vocabulary but also in phonology, morphology, syntax, semantics and pragmatics [2,3]. This structural diversity [4] can be coded using a set of *typological features*, including, for example, the number of consonants, the use of voice pitch to convey linguistic information (*tone*; [5]) or the canonical order of subject and verb, all of which take specific values in each language. The relationship between structural variation and the so-called universals of language, and the related issues concerning the nature of the constraints governing this diversity, are hotly debated issues [6], but it is clear that both cultural evolutionary processes akin to those acting on biological systems, and factors pertaining to human perception, articulation, cognition and sociality play a major role [7,8].

It is generally accepted that our capacity for speech and language rests on species-specific genetic factors, but it is currently unclear how these might be language-specific [9]. At the other end of the spectrum, it is also overwhelmingly clear that individual variation in language and speech, both normal and pathological, has strong genetic components, showing moderate to large heritabilities and confirmed by the recent characterization of various genes [10,11]. However, the possible influence of population-level genetic diversity on linguistic structural variation has not been systematically considered until the recent proposal that the distribution of *linguistic*

tone is influenced by the population frequency of the derived haplogroups of *ASPM* and *Microcephalin*, two genes involved in brain growth and development [12]. This influence is hypothesized to be mediated by a genetic bias in the acquisition and/or processing of linguistic tone—a bias that is weak at the individual level but amplified by the cultural transmission of language in populations across generations [13]. Such a proposal is supported by several computational and mathematical models showing that small biases can indeed be made manifest by cultural processes [14,15], by experiments where chains of adult humans learning artificial languages produce strong systematicity [16,17], and by zebra finches recovering the species-specific song through iterated learning across generations [18].

In general, cultural processes operate on shorter time-scales than genetic ones, with cultural change out-pacing genetic change [19,20], but this does not preclude the existence of extremely stable aspects of language [20,21]. This cultural stability can be owing to several factors, such as strong constraints generated by ecological constants, structured cultural systems with high-connectivity 'core' components and the frequency of use of parts of language [20,22]. The *genetic-biasing hypothesis* [12] thus proposes that the 'genetic anchoring' of cultural traits is yet another factor influencing their stability.

Linguistic *tone* is affected by regular processes of language change and contact: it can be gained or lost, it can complexify or simplify and it can be borrowed across language boundaries [5,23–25]. From a purely linguistic point of view, *tone* is just another aspect of language [5], and there is no *a priori* linguistic reason to expect that it would be very stable [26]. However, if linguistic *tone* is indeed under genetic biasing, then it is

expected that its dynamics would tend to correlate with that of the biasing genes. This, in turn, would result in *tone* being more resistant to 'regular' language change and more stable than other linguistic features. Unfortunately, the currently available data do not allow a more precise identification of the locus of this stability, which could result from various aspects of tonogenesis and/or tone loss. Another consequence of this hypothesis is that tone (and non-tone) languages should cluster geographically and be affected by contact and language shift, as the bias would favour both language-internal and contact-induced changes in the direction of the bias. However, this aspect is not explicitly tested in the present paper, but these areal effects of the biasing genes should still be manifest as a 'genetic anchoring' of tone. Future computational work must investigate the capacity of such phylogenetic methods to detect various types of stability owing to, for example, transmission across language shifts or slow change through vertical transmission.

Here this prediction concerning the stability of linguistic *tone* is tested using a phylogenetic approach inspired from biology [27]. Such phylogenetic methods have been successfully applied to language data [28] and are beginning to be widely accepted as an appropriate methodology for certain types of linguistic questions. However, some unclear issues do persist, especially the adequacy of trees to model aspects of language history given the known effects of borrowing and language shift (highly similar to the issues of *Horizontal Gene Transfer*; [29]), the appropriateness of models of character change and the rooting of the language trees. Given these and several other issues specific to language [22,28], I have adopted a general strategy of using several methods, datasets, codings and data-analysis strategies, as follows.

First, the main interest is in comparing the stability of *tone* with the stability of as many linguistic features as possible. The World Atlas of Language Structures (WALS) database [3] is currently the most comprehensive typological resource available, covering 141 features and 2650 languages, but owing to high levels of missing data (electronic supplementary material, the datasets), I have selected a more limited but still comprehensive subset of this database. To control for the effects of the *coding* of linguistic features, I have generated two datasets: the original *polymorphic* coding, where each feature has a specific number of values, and a linguistically informed *binary* recoding which, for some features, resulted in two related but distinct 'aspects'. For example, *tone* as a polymorphic feature has three possible values (*No tones*, *Simple tone system* and *Complex tone system*), which can be collapsed meaningfully into two binary 'aspects': *tone1* (*No tones* versus any type of tone system) and *tone2* (*Complex tone system* versus anything else; electronic supplementary material, linguistic features and table S1).

Second, the hypothesis concerns *global* tendencies, averaging across the whole world and across different language families. Generally, historical linguists consider attempts at grouping together accepted language families into 'macro-families' [30–32] as very problematic [33–35]. Therefore, I have treated language families as *independent phylogenies* on which inferences of the rates of change have resulted in a set of *posterior distributions* of rates, one distribution per language family, for each

linguistic feature. However, given the differences among historical linguists concerning the actual details of various language families, I have decided to use the two most comprehensive databases of such classifications currently available: the WALS [3] and the Ethnologue ([1]; please note that these classifications are not independent: electronic supplementary material, linguistic classifications). In total, 41 language families have been used (electronic supplementary material, the datasets): Afro-Asiatic, Algic, Altaic, Arawakan, Australian, Austro-Asiatic, Austronesian, Aymaran, Cariban, Chibchan, Chukotko-Kamchatkan, Dravidian, Eskimo-Aleut, Hokan, Indo-European, Iroquoian, Khoisan, Macro-Ge, Mataco-Guaicuru, Mayan, Na-Dene, Nakh-Daghestanian, Niger-Congo, Nilo-Saharan, North Caucasian, Oto-Manguean, Penutian, Salishan, Sepik, Sino-Tibetan, Tacanan, Tai-Kadai, Trans-New Guinea, Tucanoan, Tupian, Uralic, Uto-Aztecan, Wakashan, West Papuan, Yanomam, and Yukaghir.

Third, to address possible artefacts introduced by the specific character model and rate estimation, I have used two Bayesian phylogenetic software packages: the widespread MRBAYES 3 [36] and the custom-written BAYESLANG. In both cases, the historical linguistic classification (either from WALS or Ethnologue) is taken as given and used to compute the rates of change of the linguistic features considered, but BAYESLANG explicitly considers it as rooted and deals with the unresolved nodes (polytomies) as such, while MRBAYES needs an *outgroup* in order to root the tree and attempts to fully resolve it as well. For MRBAYES, the linguistic classification was transformed into *constraints* on the admissible topologies and the binary data were considered as *restriction*, while the polymorphic as *standard* [36]. BAYESLANG was run for 5 000 000 generations (1 000 000 burn-in, one cold and six heated chains) and MRBAYES was run for 5 000 000 generations (1000 sampling frequency, 1000 generations burn-in) and all runs have converged (log-likelihood plots and the potential scale reduction factor; [36]). The methods used to estimate the rates differ markedly between programs, with BAYESLANG estimating the minimum number of changes from the inferred ancestral to the observed states, akin to a parsimony approach and producing an underestimate of the actual number of changes, while MRBAYES estimated the rates of change using a gamma model (electronic supplementary material, stability estimation).

Finally, MRBAYES requires an *outgroup* to root the phylogeny and there are many issues surrounding the choice of outgroups in general [37] and in linguistics in particular, owing mainly to the contentious issue of the above family-level relationships, compounded by non-vertical processes in language. To systematically investigate the potential influence the choice of outgroup might have on the rates estimated by MRBAYES, I have selected 23 typologically and geographically very diverse language isolates to be used as outgroups (electronic supplementary material, outgroups). For the Ethnologue classification, I have used each of these isolates as the outgroup, therefore replicating 23 times each run for each relevant combination of parameters (see below). Owing to computational constraints and the very high correlation between the two classifications (see below), for the WALS classification I have used only two of these language

isolates, namely the geographically and typologically distinct *Ainu* and *Basque*.

In total, I conducted 54 runs using multiple data codings, linguistic classifications, software implementations and outgroup choices (electronic supplementary material, the datasets), allowing the quantification and minimization of potential artefactual results. To ensure the comparability across such a diverse range of results, I converted the rate estimations for each linguistic feature into *ranks*, ranging from the *most stable* to the *most unstable*, resulting in posterior distributions of feature stability ranks. The analyses presented below take these ranks as their input.

2. RESULTS

(a) *Summary properties of the feature-rank posterior distributions*

Family-level distributions (not shown) are unimodal for most features and most language families, although several cases of bi- or multi-modality suggest problems with the data, the inference parameters and/or the adequacy of a tree model in these cases. The *macro-area*- and *world*-level combined distributions generally tend to feature multiple peaks, strongly suggesting different language family- and macro-area-specific processes. Nevertheless, there seem to exist systematic differences between features in their stability patterns.

Owing to these strong deviations from normality, the posterior rank distributions are summarized by both *means* and *medians*. However, independent of other parameters, the *means* and *medians* agree very strongly ($0.94 \leq r \leq 0.98$, $p < 10^{-6}$), suggesting that the results are not artefacts of the data compression strategy.

(b) *Effects of linguistic classification and outgroup choice*

The two historical linguistic classifications (WALS and Ethnologue) produce very similar results across datasets ($0.96 \leq r \leq 0.99$, $p < 10^{-10}$), suggesting that the differences in stability between typological features are not artefacts of these particular historical classifications.

The effects of the outgroup choice on the rates estimated by MRBAYES seem to be minimal given that the feature rankings produced using *Basque* and *Ainu* as outgroups correlate very strongly ($0.82 \leq r \leq 0.86$, $p < 10^{-10}$), independently of data coding and linguistic classification. This is reinforced by using all 23 language isolates (including *Basque* and *Ainu*) as outgroups for the Ethnologue classification and both codings, resulting in highly similar ranks: the first principal component across outgroups, PC_1, explains 79 per cent of the variance, and the Pearson correlation ranges between $0.49 \leq r \leq 0.92$, $p < 10^{-6}$ with a mean $r = 0.78$. Given these high similarities, the posterior rank distributions produced by the different outgroups were combined and an 'outgroup-average' posterior distribution was extracted for further analysis.

(c) *Binary coding*

The following four datasets have been analysed: MRBAYES with the WALS classification combining *Ainu* and *Basque* as outgroups, MRBAYES with the Ethnologue classification combining all 23 language isolates as outgroups and

BAYESLANG using both classifications. The feature stability ranks produced by these datasets agree strongly (mean $r = 0.78$, range $0.59 \leq r \leq 0.98$, $p < 10^{-8}$; electronic supplementary material, table S2), confirmed by a principal components analysis, where PC_1 explains 81.4 per cent of the variance and represents the commonality between datasets (electronic supplementary material, table S3). Moreover, the average feature rank across these datasets correlates very strongly with PC_1 ($r = 0.99$, $p < 2.2 \times 10^{-16}$) and classifies *tone2* (complex tone systems; electronic supplementary material, linguistic features and table S1) as the 8th (out of 86), with *tone1* (simple tone systems) as the 23rd (out of 86; electronic supplementary material, table S4).

(d) *Polymorphic coding*

There are four datasets mirroring the binary case and, as above, the stabilities agree strongly across datasets (mean $r = 0.71$, range $0.51 \leq r \leq 0.99$, $p < 10^{-5}$; electronic supplementary material, table S5), with the first component PC_1—the commonality—explaining 76.1 per cent of the variance (electronic supplementary material, table S6). Again, the average rank correlates very strongly with PC_1 ($r = 0.99$, $p < 2.2 \times 10^{-16}$) and classifies *tone* as the eighth most stable out of 68 polymorphic features (electronic supplementary material, table S7).

(e) *The relationship between polymorphic and binary ranks*

For any single polymorphic feature there can be more than one corresponding binary feature, capturing different linguistically relevant aspects (electronic supplementary material, linguistic features and table S1) and potentially having different stabilities (electronic supplementary material, tables S4 and S7). Because of this, the agreement between data codings varies dramatically across datasets and methods ($0.18 \leq r \leq 0.99$; electronic supplementary material, table S8), but tends to be rather strong (mean $r = 0.61$, median $p = 6.5 \times 10^{-9}$). PC_1 (representing the agreement between rankings) explains 67.4 per cent of the variance, and PC_2, explaining 16.1 per cent, contrasts the binary and polymorphic codings (electronic supplementary material, table S9). The average rank across both codings correlates very strongly with PC_1 ($r = 0.99$, $p < 10^{-15}$), while the average binary and average polymorphic ranks correlate moderately with one another ($r = 0.58$, $p = 7.1 \times 10^{-8}$).

The complex relationship between polymorphic and corresponding binary features is shown in figure 1 (see also electronic supplementary material, table S10). *Tone* is very stable as a polymorphic feature (figure 1*b*; one-sample *t*-test: $t_{56} = 9.7$, $p = 1.35 \times 10^{-13}$) and *tone2* is very stable both as a binary feature (figure 1*b*; $t_{70} = 12.04$, $p < 2.2 \times 10^{-16}$) and overall (figure 1*a*; $t_{70} = 12.27$, $p < 2.2 \times 10^{-16}$), with *tone1* relatively stable in both comparisons (figure 1*a*; $t_{70} = 4.35$, $p = 6.7 \times 10^{-9}$; figure 1*b*; $t_{70} = 4.35$, $p = 4.5 \times 10^{-5}$). Compared only with *phonological* features (electronic supplementary material, table S10), *tone* is stable as a polymorphic feature (5th out of 13; $t_{12} = 2.4$, $p = 0.034$), *tone2* is very stable both as a binary feature (3rd out of 19; $t_{18} = 5.1$, $p = 7.65 \times 10^{-5}$) and overall (3rd out of 19; $t_{18} = 4.73$, $p = 0.00017$), while *tone1* is of average stability as a
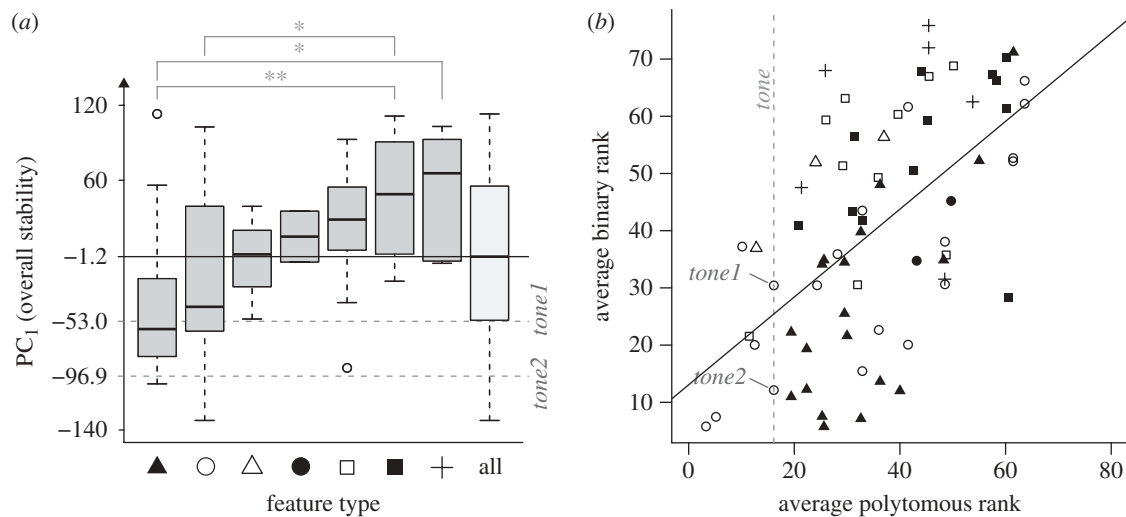
Figure 1. The stability of typological features. (*a*) Boxplot of scores of binary features on PC$_1$ (overall stability ranging from the most stable on the bottom to the most unstable on top) by feature type (sorted by increasing median stability), and significant differences after Tukey's HSD. (*b*) Scatterplot of binary features on the average rank of the corresponding polymorphic feature versus the average rank of the binary features, with their regression line. Black triangle, features of type Word Order; white circle, Phonology; white triangle, Nominal Syntax; black circle, Morphology; white square, Verbal Categories; black square, Nominal Categories; cross, Simple Clauses; and 'all', all types of features combined.

binary feature (8th out of 19; $t_{18} = 0.82$, $p = 0.42$) and marginally stable overall (7th out of 19; $t_{18} = 2.1$, $p = 0.052$).

(f) *The stability of types of feature*

The WALS classifies the linguistic features into various categories and the issue of the relative stabilities of such categories is of appreciable interest to historical linguists and typologists [38,39]. Using these methods, there do not seem to be any significant differences between the stabilities of types of *polymorphic* features (one-way ANOVA: $F_{6,64} = 1.55$, $p = 0.18$), but types of *binary* features do differ ($F_{6,64} = 6.05$, $p = 4.7 \times 10^{-5}$; figure 1*b*), with *Word Order* on average more stable than *Nominal Categories* and *Simple Clauses*, and *Phonology* more stable than *Nominal Categories* (after Tukey's HSD multiple comparisons correction; [40]).

3. DISCUSSION AND CONCLUSIONS

The work reported here represents, to my knowledge, the first investigation of the stability of structural properties of language using a phylogenetic perspective across a large set of language families and geographical areas, complementing the more focused study of Greenhill *et al.* [22]. This approach takes into account language family-level uncertainty in the inferred parameters through the use of Bayesian phylogenetic methods, that generate whole posterior distributions for these parameters instead of point estimates [27,36]. Moreover, the specific design introduced in this paper attempts to control for other potential sources of artefacts, such as the inference and rate-estimation *algorithms*, the *linguistic classification* needed to guide the inference, the *coding* of the data and the choice of *outgroup* required to root the phylogeny. Therefore, 54 separate datasets have been analysed, each comprising a number of language families treated as independent phylogenies, resulting in the analysis of 113 246

phylogenies, each generating a set of posterior distributions of linguistic features' stability ranks.

The results show that, overall, these different methods, data codings, linguistic classifications and outgroup choices largely agree on the inferred ranking of linguistic features from the most to the least stable. This agreement suggests that, in this case at least, despite the various issues concerning the current application of phylogenetic methods to typological data, they produce *reliable* results.

Linguistic tone seems stable relative to a large set of features covering many aspects of language, both as a polymorphic feature and in its two binary aspects (being a tone language or not—*tone1*—and having a complex tone system—*tone2*), supporting one of the predictions of the *genetic-biasing hypothesis* [12]. However, the *validity* of these results must be taken as suggestive, given the issues with the primary data (low coverage, coding decisions and chance similarity owing to the restricted range of typological features), linguistic classifications considered, the usage of language isolates as outgroups and the liability of typological features to areal effects. In this vein, Greenhill *et al.* [22] have found that typological features might be less appropriate than basic vocabulary to the application of phylogenetic methods and that these two aspects seem to evolve at comparable rates.

A more specific test of the genetic-biasing hypothesis would be to estimate the *correlated evolution* [41,42] between *tone* and the population frequency of the two biasing alleles, but this requires much better linguistic and genetic data and stronger assumptions concerning the adequacy of trees as models of linguistic and genetic relationships between populations. Another direction, currently underway, aims at operationalizing inter-individual variation in the acquisition and/or processing of linguistic *tone* and estimating the *heritability* of this variance [43] and its *association* with the two biasing

genes [44]. Recently, encouraging results suggest that common polymorphisms in these two genes are associated with normal variation in brain morphology [45,46], while Christiansen *et al.* [47] report an association between polymorphisms of *ASPM* and various language measures.

The method introduced in this paper suggests some intriguing patterns: *FrRoundV* (presence of Front Rounded Vowels) is extremely stable (2nd as polymorphic, electronic supplementary material, table S7 and 4th as binary, electronic supplementary material, table S4) and has a skewed geographical distribution [48]. Likewise, *MTPron* (personal pronouns patterning as *m* in the 1st person and a coronal obstruent in the 2nd) and *NMPron* (*n* in the 1st person and *m* in the 2nd; [49]) are relatively stable (14th and 9th as polymorphic features, respectively), apparently supporting Nichols' claim [50] that these are indicators of very old demographic and linguistic processes. The reasons for the stability of such features warrants future investigation, as some might represent new cases of genetic biasing.

This paper has specifically focused on the global stability of linguistic features, but the variation observed within and between language families [22] and geographical areas requires further research as it might offer clues to language family- and area-specific processes. Especially interesting in this respect are the cases of multimodal posterior distribution of rates and the possible existence of concordances within geographical areas, which could point to large-scale contact phenomena and/or ancient language expansions.

REFERENCES

1 Lewis, M. (ed.) 2009 *Ethnologue: languages of the world*, 16th edn. Dallas, TX: SIL International.

2 Croft, W. 2002 *Typology and universals*. Cambridge, UK: Cambridge University Press.

3 Haspelmath, M., Dryer, M., Gil, D. & Comrie, B. (eds) 2005 *The world atlas of language structures*. Oxford, UK: Oxford University Press.

4 Nettle, D. 1999 *Linguistic diversity*. Oxford, UK: Oxford University Press.

5 Yip, M. 2002 *Tone*. Cambridge, UK: Cambridge University Press.

6 Evans, N. & Levinson, S. 2009 The myth of language universals. *Behav. Brain Sci.* **32**, 429–448. (doi:10.1017/S0140525X0999094X)

7 Christiansen, M. & Chater, N. 2008 Language as shaped by the brain. *Behav. Brain Sci.* **31**, 489–509. (doi:10.1017/S0140525X08004998)

8 Pagel, M. 2009 Human language as a culturally transmitted replicator. *Nat. Rev. Genet.* **10**, 405–415. (doi:10.1038/nrg2560)

9 Hauser, D. M., Chomsky, N. & Fitch, W. T. 2002 The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)

10 Bishop, D. V. M. 2009 Genes, cognition, and communication: insights from neurodevelopmental disorders. *Ann. NY Acad. Sci.* **1156**, 1–18. (doi:10.1111/j.1749-6632.2009.04419.x)

11 Fisher, S. & Scharff, C. 2009 *FOXP2* as a molecular window into speech and language. *Trends Genet.* **25**, 166–177. (doi:10.1016/j.tig.2009.03.002)

12 Dediu, D. & Ladd, D. R. 2007 Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, *ASPM* and *Microcephalin*. *Proc. Natl Acad. Sci. USA* **104**, 10 944–10 949. (doi:10.1073/pnas.0610848104)

13 Ladd, D. R., Dediu, D. & Kinsella, A. 2008 Languages and genes: reflections on biolinguistics and the nature-nurture question. *Biolinguistics* **2**, 114–126.

14 Dediu, D. 2008 The role of genetic biases in shaping language-genes correlations. *J. Theor. Biol.* **254**, 400–407. (doi:10.1016/j.jtbi.2008.05.028)

15 Kirby, S., Dowman, M. & Griffiths, T. L. 2007 Innateness and culture in the evolution of language. *Proc. Natl Acad. Sci. USA* **104**, 5241–5245. (doi:10.1073/pnas.0608222104)

16 Kirby, S., Cornish, H. & Smith, K. 2008 Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proc. Natl Acad. Sci. USA* **105**, 10 681–10 686. (doi:10.1073/pnas.0707835105)

17 Smith, K. & Wonnacott, E. 2010 Eliminating unpredictable variation through iterated learning. *Cognition* **116**, 444–449. (doi:10.1016/j.cognition.2010.06.004)

18 Fehér, O., Wang, H., Saar, S., Mitra, P. P. & Tchernichovski, O. 2009 *De novo* establishment of wild-type song culture in the zebra finch. *Nature* **459**, 564–568. (doi:10.1038/nature07994)

19 Boyd, R. & Richerson, P. 1985 *Culture and the evolutionary process*. Chicago, IL: University of Chicago Press.

20 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)

21 Hunley, K., Dunn, M., Lindström, E., Reesink, G., Terrill, A., Healy, M. E., Koki, G., Friedlaender, F. R. & Friedlaender, J. S. 2008 Genetic and linguistic coevolution in Northern Island Melanesia. *PLoS Genet.* **4**, e1000239. (doi:10.1371/journal.pgen.1000239)

22 Greenhill, S. J., Atkinson, Q. D., Meade, A. & Gray, R. D. 2010 The shape and tempo of language evolution. *Proc. R. Soc. B* **277**, 2443–2450. (doi:10.1098/rspb.2010.0051)

23 Hyman, L. M. 1978 Historical tonology. In *Tone: a linguistic survey* (ed. V. A. Fromkin), pp. 257–270. London, UK: Academic Press.

24 Svantesson, J.-O. & House, D. 2006 Tone production, tone perception and Kammu tonogenesis. *Phonology* **23**, 309–333. (doi:10.1017/S0952675706000923)

25 Hyslop, G. 2009 Kurtöp tone: a tonogenetic case study. *Lingua* **119**, 827–845. (doi:10.1016/j.lingua.2007.11.012)

26 Bowles, J. 2008 Some questions about determining causal inference and criteria for evidence. *Biolinguistics* **2**, 247–255.

27 Felsenstein, J. 2004 *Inferring phylogenies*. Sunderland, MA: Sinauer.

28 Gray, R. D., Greenhill, S. J. & Ross, R. M. 2007 The pleasures and perils of Darwinizing culture (with phylogenies). *Biol. Theor.* **2**, 360–375.

29 Doolittle, W. F. 2009 The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them. *Phil. Trans. R. Soc. B* **364**, 2221–2228. (doi:10.1098/rstb.2009.0032)

30 Cavalli-Sforza, L. L. 2001 *Genes, peoples and languages*. Berkeley, CA: University of California Press.

31 Greenberg, J. 1998 The convergence of Eurasiatic and Nostratic. In *Nostratic: sifting the evidence*

(eds J. C. Salmons & B. D. Joseph), pp. 51–60. Amsterdam, The Netherlands: John Benjamins Publishing Co.

32 Ruhlen, M. 1991 *A guide to the world's languages. Vol. I. Classification (with a postscript on recent developments).* Stanford, CA: Stanford University Press.

33 Bolnick, D., Shook, B., Campbell, L. & Goddard, I. 2004 Problematic use of Greenberg's linguistic classification of the Americas in studies of Native American genetic variation. *Am. J. Hum. Genet.* **75**, 519–523. (doi:10.1086/423452)

34 Sims-Williams, P. 1998 Genetics, linguistics, and prehistory: thinking big and thinking straight. *Antiquity* **72**, 505–527.

35 Campbell, L. 1999 Nostratic and linguistic palaeontology in methodological perspective. In *Nostratic: examining a linguistic macrofamily* (eds C. Renfrew & D. Nettle), pp. 179–230. Cambridge, UK: The McDonald Institute for Archaeological Research and Cambridge University Press.

36 Ronquist, F. & Huelsenbeck, J. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

37 Sanderson, M. J. & Shaffer, H. B. 2002 Troubleshooting molecular phylogenetic analyses. *Ann. Rev. Ecol. Syst.* **33**, 49–72. (doi:10.1146/annurev.ecolsys.33.010802. 150509)

38 Greenberg, J. 1978 How does a language acquire gender markers? In *Universals of human language, vol. III. Word structure* (eds J. Greenberg, C. Ferguson & E. Moravcsik), pp. 47–82. Stanford, CA: Stanford University Press.

39 Nichols, J. 1995 Diachronically stable structural features. In *Historical Linguistics 1993. Selected Papers from the 11th Int. Conf. on Historical Linguistics*, Los Angeles, 16–20 August 1993 (ed. A. Henning), pp. 337–355. Amsterdam, The Netherlands: John Benjamin.

40 Keselman, H. J. & Toothaker, L. E. 1974 Comparison of Tukey's T-method and Scheffé's S-method for various

numbers of all possible differences of averages contrasts under violation of assumptions. *Educ. Psych. Meas.* **3**, 511–519.

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

42 Felsenstein, J. 1985 Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15.

43 Lynch, M. & Walsh, B. 1998 *Genetics and analysis of quantitative traits.* Sunderland, MA: Sinauer Assoc. Inc.

44 Neale, B., Ferreira, M., Medland, S. & Posthuma, D. (eds) 2007 *Statistical genetics: gene mapping through linkage and association.* New York, NY: Taylor & Francis.

45 Rimol, L. M. *et al.* & the Alzheimer's Disease Neuroimaging Initiative 2010 Sex-dependent association of common variants of microcephaly genes with brain structure. *Proc. Natl Acad. Sci. USA* **107**, 384–388. (doi:10. 1073/pnas.0908454107)

46 Wang, J., Li, Y. & Su, B. 2008 A common SNP of MCPH1 is associated with cranial volume variation in Chinese population. *Hum. Mol. Genet.* **17**, 1329–1335. (doi:10.1093/hmg/ddn021)

47 Christiansen, M. H., Kelsey, K. & Tomblin, J. B. 2008 Associations of ASPM genetic markers with language and non-verbal intelligence. In *33rd Ann. Boston Univ. Conf. Lg. Develop.*, Boston, MA.

48 Maddieson, I. 2008 Front rounded vowels. In *The world atlas of language structures online* (eds M. Haspelmath, M. S. Dryer, D. Gil & B. Comrie). Munich, Germany: Max Planck Digital Library. See http://wals.info/feature/11.

49 Nichols, J. & Peterson, D. A. 2008 N-M pronouns. In *The world atlas of language structures online* (eds M. Haspelmath, M. S. Dryer, D. Gil & B. Comrie). Munich, Germany: Max Planck Digital Library. See http://wals.info/feature/137.

50 Nichols, J. 1992 *Linguistic diversity in space and time.* Chicago, IL: Chicago University Press.