BRILL
Language Dynamics and Change 1 (2011) 205–240
brill.nl/ldc

Correlates of Reticulation in Linguistic Phylogenies

Søren Wichmann,[a] Eric W. Holman,[b] Taraka Rama[c] and Robert S. Walker[d]

a) Max Planck Institute for Evolutionary Anthropology
wichmann@eva.mpg.de

b) University California, Los Angeles
holman@psych.ucla.edu

c) University of Gothenburg & Max Planck Institute for Evolutionary Anthropology
tarakaramadaiict@gmail.com

d) University of Missouri
walkerro@missouri.edu

Abstract

This paper discusses phylogenetic reticulation using linguistic data from the Automated Similarity Judgment Program or ASJP (Holman et al., 2008; Wichmann et al., 2010a). It contributes methodologically to the examination of two measures of reticulation in distance-based phylogenetic data, specifically the δ score of Holland et al. (2002) and the more recent *Q*-residuals of Gray et al. (2010). It is shown that the δ score is a more adequate measure of reticulation. Our empirical analyses examine possible correlations between δ and (a) the size (number of languages), (b) age, and (c) heterogeneity of language groups, (d) linguistic isolation of individual languages within their respective phylogenies, and (e) the status of speech forms as dialects or recently emerged languages. Among these, only (d) is significantly correlated with δ. Our interpretation is that δ is a realistic measure of reticulation and sensitive to effects of socio-historical events such as language extinction. Finally, we correlate average δ scores for different language families with the goodness of fit between ASJP and expert classifications, showing that the δ scores explain much of the variance.

Keywords

phylogenetics; delta; *Q*-residuals; reticulation; historical linguistics; ASJP

1. Introduction

In this paper we are interested in the causes and consequences of phylogenetic reticulation. Reticulation within phylogenies is known to occur among

 DOI: 10.1163/221058212X648072

biological lineages and can be defined as the amalgamation of previously separate branches of a phylogenetic tree. In biology, reticulation is usually produced by horizontal gene transfer and by hybridization. Moreover, reticulation can be induced in estimated phylogenies by several additional factors including reversal, convergent evolution (homoplasy), and coding or sampling errors. Some but not all of these phenomena have clear analogues in the linguistic evolution of lexical and phonological features, which are the kinds of features we are drawing upon here. Horizontal transfer (diffusion) is known to occur for all aspects of language structure, including, of course, lexical items (Haspelmath and Tadmor, 2009). Even the lexical items pertaining to the well-known Swadesh list, which are often considered to not be frequently borrowed, have a borrowing rate exceeding 8 % in at least one large sample of languages (Holman et al., 2008). The lexical analogue to reversal would be the loss and subsequent reappearance of a vocabulary item, a phenomenon which is at best marginal. Homoplasy most frequently occurs as similar but independent phonological changes. These may pose difficulties for the historical linguist, who often has to refer to experience or intuition in order to decide whether such changes are in fact independent or whether they are diffused or shared via a common ancestor. Finally, large linguistic databases are prone to coding errors, references to erratic classifications, and misidentification of certain languages, and they often lack some data.

All these problems are potential sources of error that will induce conflicts in phylogenetic signals. For these various reasons we should be surprised if all linguistic phylogenies were completely treelike, and, in fact, they rarely are. Among the various causal factors mentioned, homoplasy, especially, is expected to contribute its portion of noise, but its contribution would be difficult to assess, so we are not going to investigate this factor here. Among potential problems with data, those relating to errors cannot be estimated because the amount of error is unknown. Missing data points, however, are readily identified and their effect will be studied here. Our primary focus is on other possible factors that might affect measures of reticulation in linguistic phylogenies—factors that are directly interesting from the linguist's point of view. Thus, we are going to look at whether the size, age or heterogeneity of a language group influence reticulation; whether dialect chains or recently emerged languages contain more reticulation; and whether isolates within families (i.e., languages having no closer relatives than the protolanguage itself) are more or less reticulate than languages having diverged from some intermediate language. Finally, we investigate how more versus less reticulate networks produced through a single, consistent method of classification compare with the views of experts working within traditional historical linguistic frameworks.

Phylogenetic networks (Huson et al., 2010) represent a useful visual approach to the discovery of reticulation. Such networks are increasingly used as tools for representing historical relationships among languages even if they are essentially phenetic. That is, they do not actually embody phylogenetic models as such, but are simply tools for visualizing degrees of similarity among languages. Pioneering papers in this area include Forster et al. (1998), Forster and Toth (2003), and Bryant et al. (2005). The more recent widespread use of networks is illustrated in papers on a variety of historical linguistic topics, including the status of creoles (Bakker et al., 2011), the use of structural features to assess historical relationships (Dunn et al., 2005; Wichmann and Saunders, 2007), and the classification of individual language groups such as Indo-European (Gray et al., 2010), Bantu (Holden and Gray, 2006), Quechuan/Aymaran (McMahon et al., 2005), Karnic languages of Australia (Bowern, 2010), and varieties of English (McMahon et al., 2007; Heggarty et al., 2010; Wichmann and Urban, in press), to name but a few. This recent popularity is in no small measure due to readily available software to produce such networks, primarily SplitsTree (Huson and Bryant, 2006).

Since networks reveal contradictory phylogenetic signals, they are useful for comparing a situation with much reticulation to one where treelikeness predominates. But, beyond the extreme and obvious situations, networks can pose difficulties for pure, visual interpretation because of the large amount of information they contain. Thus, some non-visual, quantitative way of expressing reticulation associated with taxa, clades or entire families is necessary for moving away from the mainly aesthetic impression to a more principled interpretation of phylogenetic networks. We therefore discuss methods of actually measuring reticulation. Two different reticulation metrics will be treated in some detail, since, with the notable exception of Gray et al. (2010), such metrics have not yet been applied to linguistic data. Moreover, one of the metrics—the one called *Q*-residual—has not hitherto been fully described or extensively investigated for its properties.

The empirical data drawn upon in this paper are from the ASJP project. Under the auspices of this project, 40-item word lists have been collected for well over one half of the world's languages. Version 14 of the ASJP database (Wichmann et al., 2011) is used in all of the present analyses except for one described in Section 4. The word lists are compared by applying a modified version of the Levenshtein distance called LDND (Levenshtein Distance Normalized Divided). To calculate LDND between a pair of lists,

1. the Levenshtein (or edit) distance, LD, is found between words for the same item in the different lists;

2. LD is normalized for word length by dividing the LD by the number of symbols in the longer of the two strings compared, yielding LDN (Levenshtein Distance Normalized);
3. LDN is averaged across items and corrected for chance similarity by dividing the average LDN for words referring to the same concept by the average LDN for word pairs whose members refer to different concepts, yielding LDND.

LDND is described in several papers, including Wichmann et al. (2010a), where the measure is also tested and discussed in more detail. In the full ASJP database some languages are represented by several lists from different sources or dialects; except in the one analysis in Section 4, a single list is arbitrarily chosen when more than one is available, such that all the lists are identified by different ISO 639-3 codes. Known loanwords are identified in some of the ASJP lists; these are retained in the present analyses to preserve a complete picture of diffusion.

Since reticulation reflects deviations from a phylogenetic tree, measures of reticulation presume taxa that are phylogenetically related to each other. Although language families are defined as groups of phylogenetically related languages, their status as true genealogical units remains uncertain for several of the families in compilations such as *Ethnologue* (Lewis, 2009) and the *World Atlas of Language Structures* (Haspelmath et al., 2005; henceforth *WALS*). Examples include large families such as Niger-Congo, Nilo-Saharan, Altaic, and Australian. We try to avoid spurious effects on reticulation induced by the lack of phylogenetic relationship among some of the languages in some families. Consequently, all the analyses except one in Section 4 will draw upon a more recent and conservative classification by Harald Hammarström (personal communication, 2011). His classification is outlined in the online appendix to Hammarström (2010), and each of the languages in the ASJP database, starting from Version 14 (Wichmann et al., 2011), is classified according to Hammarström's scheme (which, especially with regard to subclassification of families, is expected to be modified in the future).

2. Reticulation Metrics

The two reticulation metrics we will review here take distances as input rather than characters (see Wichmann, 2010 for an introduction to this difference). Because our data are distance-based, we exclude from consideration measures of treelikeness that operate on data encoded as characters, such as Pagel's lambda (Pagel, 1999), or related methods of assessing the validity of a phylogeny, such as the bootstrap (Felsenstein, 1985) or Bayesian posterior probabilities.

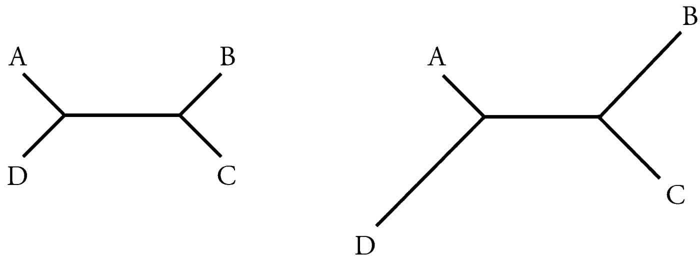

Figure 1a–b. Two examples of additive trees.

Both of the metrics to be investigated, δ and *Q*-residuals, measure reticulation at the level of quartets. A quartet is a set of four taxa pertaining to a given phylogeny. If the matrix of distances among the four taxa can be represented by a tree, such that the distances along the branches (edges) connecting the four languages precisely represent the distances in the matrix, then the matrix is said to be *additive*. Two examples of additive trees are given in Fig. 1.

An additive tree satisfies the so-called four-point condition on distances between taxa. We use the notation |…| to indicate the distance between two taxa. There are three ways to partition a quartet into two pairs of taxa: (A, B) and (C, D), (A, C) and (B, D), and (A, D) and (B, C). Each partition corresponds to a sum of two pairwise distances: |AB| + |CD|, |AC| + |BD|, and |AD| + |BC|. The three sums can always be ordered from largest to smallest, for instance:

(1) $|AC| + |BD| \geq |AB| + |CD| \geq |AD| + |BC|$

The four-point condition states that the two largest sums must be equal; in other words, if (1) holds, then:

(2) $|AC| + |BD| = |AB| + |CD| \geq |AD| + |BC|$

The two additive trees in Fig. 1 do in fact satisfy (2). In general, if all the quartets satisfy the four-point condition, then an additive tree will predict all the pairwise distances. If we want to measure the deviation from treelikeness of a quartet we can use the deviation from (2) as our metric. The reticulation measure can be extended to express how well a given taxon fits within a larger network by averaging the measure for all the quartets in which the taxon participates (Holland et al., 2002), or it can be extended to express how treelike a larger set of taxa is by averaging the measure over all quartets in the set.

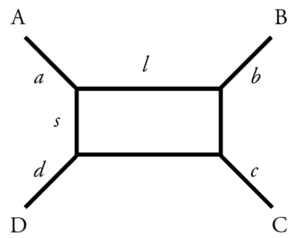

Figure 2. Representation of a reticulate quartet.

In the next two sections we will look individually at the two different ways of deriving reticulation measures from deviations from the four-point condition that have so far been proposed.

2.1. δ

Holland et al. (2002) propose a measure of reticulation called δ. They point out that the pairwise distances in a quartet can always be represented as in Fig. 2. If we consider the reticulate quartet in Fig. 2, their measure can simply be expressed as s/l, the shorter side of the box divided by the longer one.

While this is what the measure reduces to, it takes as its point of departure the four-point condition. First, the difference between the largest and the next largest sum of distances is found (i.e., the sums that should have been equal if the four-point condition held), which is $(|AC| + |BD|) - (|AB| + |CD|)$. This is normalized (divided) by $(|AC| + |BD|) - (|AD| + |BC|)$, the largest sum minus the smallest sum. Seeing that this is the same as s/l is easy by substituting the symbols for the lengths (weights) of each edge into the expression. The values of δ range from 0 (where the quartet is additive) to 1 (where $s = l$). It is worth stressing the fact that the weights of the terminal branches *a*, *b*, *c*, and *d* do not play a role—or rather, that they cancel out—in the calculation of δ.

Holland et al. (2002) define δ in terms of distances, which can be either distances observed in data or distances predicted by a model. The SplitsTree software originally developed by Huson and Bryant (2006) fits a reticulate model like Fig. 2 to a matrix of distance data; version 4.12.3 of the software calculates δ from the distances in the model (an earlier version performed miscalculations). One of the applications, namely the δ-plots of Holland et al. (2002), is implemented in the ape package (Paradis et al., 2004) of R (R Development Core Team, 2011).

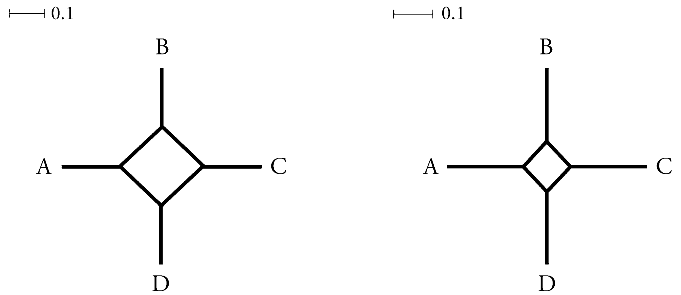

Figure 3a–b. Two quartets. The Q-r value is 0.36 in the quartet to the left and 0.14 in the one to the right. (For the purpose of this illustration, Hamming distances, indicated by the scale bars, are normalized by the total number of characters; for instance, the absolute distance |AB| = 3 in Figure 3a is normalized to 0.5).

2.2. Q-*Residual*

Q-residual (henceforth *Q*-r) is another measure of reticulation, which is briefly described by Gray et al. (2010). To calculate *Q*-r, the pairwise distances are first normalized by finding the average of all the distances within the family and then dividing each distance by the average. By analogy with squared error, *Q*-r is defined for a given quartet as the square of the difference, which should be 0 under the four-point condition; in other words, *Q*-r is $(|AC| + |BD| - |AB| - |CD|)^2$, where |AC|, |BD|, etc. are normalized distances. Like δ, *Q*-r can be calculated either from distances observed in data or from distances predicted by a model. *Q*-r is affected by the lengths of the terminal branches denoted *a*, *b*, *c*, and *d* in Fig. 2. We can verify this through some toy examples given in Fig. 3. We posit four taxa. In Fig. 3a each is defined by a unique character and, in addition, there is a character by which A and B are similar over against C and D, but also one by which A and C are similar over against B and D—the maximally reticulate situation. The characters are transformed to Hamming distances, giving 3.33 for the average distance and 0.36 for *Q*-r. We can now add additional length to the branches leading to each individual taxon by positing an additional character for each that is not shared by the others. Then the average distance increases to 5.33 and *Q*-r diminishes to 0.14.

The dependency on lengths of terminal branches renders *Q*-r a measure very different from δ. For the latter, the two quartets in Fig. 3 are exactly equally reticulate, with δ = 1 in both cases. In contrast, according to the *Q*-r measure, the quartet to the left has a reticulation value of 0.36 that is much greater than the value of 0.14 for the quartet to the right. We regard the δ approach as more adequate since the distinctiveness of one or more taxa in relation to the

Table 1. Spearman rank correlations for δ, *Q*-r, *N*, LDND, and att across families.

Entities correlated	ϱ	*p*
δ ~ *Q*-r	.5088	< .000001
δ ~ *N*	.1586	.1150
Q-r ~ *N*	-.0876	.3861
δ ~ LDND	.0777	.4422
Q-r ~ LDND	-.5214	< .0000001
δ ~ att	-.2741	.0058
Q-r ~ att	-.1617	.1081

others should not affect the measure. Since *Q*-r is a new metric, we nevertheless submit it to some further testing, using empirical data.

The SplitsTree 4.12.3 implementation of *Q*-r had an error (confirmed by David Bryant in personal communication, Jan. 16, 2012), which seems to have carried over to *Q*-r values reported in Gray et al. (2010). Values for δ and *Q*-r cited in the present paper derive from our own software implementations.

2.3. *Empirical Comparison of* δ *and* Q-*r*

In Table 1 we provide Spearman rank correlations across families among δ, *Q*-r, and three variables of primarily methodological interest: *N*, LDND, and att. *N* is the total number of languages with different ISO 639-3 codes in the family. LDND is the mean of LDND across all language pairs in the family; this indicates the lexical heterogeneity of the family, given that LDND is high between dissimilar languages and low between similar ones. Finally, att is the mean number of items attested in the lists for languages in the family. The values of δ and *Q*-r are averaged across all quartets in the family. Spearman correlations are calculated across the 100 Hammarström families with *N* at least 4, the smallest size for which δ and *Q*-r are defined. Significance testing for the Spearman correlations was carried out using Algorithm AS 89 (Best and Roberts, 1975) as implemented in R (R Development Core Team, 2011). The data on which the correlations are based are provided in Table 2 below.

The results in Table 1 show, first, that δ and *Q*-r are significantly correlated, presumably because of the common elements in their definitions. The other numbers in Table 1 are helpful for deciding which measure of reticulation is better. There is no obvious reason to expect the size or heterogeneity of a language family to influence reticulation. In fact, *N* is not correlated with either δ or *Q*-r. LDND is likewise uncorrelated with δ, but in contrast it shows a strong negative correlation with *Q*-r, so much so that *Q*-r appears to be as revealing of raw distances between taxa as anything else. The toy example in Section 2.2 (see Fig. 3) illustrates how this can come about, i.e., how greater distances among languages will lead to smaller values of *Q*-r. The significant negative correlation between δ and att shows that reticulation grows with the number of missing items in word lists, which is entirely expected from the greater random variability in lists with fewer items. Again we see a contrast with *Q*-r, which does not exhibit significance in its sensitivity to the number of attestations, further calling into question its usefulness as a measure of reticulation.

These empirical results, along with the general observations in Section 2.2 above, strongly indicate that δ is more adequate as a measure of reticulation than *Q*-r. Thus, all claims made in the following about the causes and consequences of linguistic reticulation will be based solely on δ. The influence of the number of attestations shows that for correlations between δ and other factors it would normally be necessary to take this confounding factor into account.

Table 2 shows data for the entities correlated in Table 1: δ, *Q*-r, LDND, *N*, and att for Hammarström's families. Also supplied are ages of the individual families. These will be discussed in section 3.1 below.

Table 2. Data on δ, *Q*-r, lexical distance, number of languages, attested items, and age for the families in Hammarström's classification having at least 4 languages.

Family	δ	*Q*-r	LDND	*N*	att	age
Abkhaz-Adyge	0.0305	0.000875	80.93	5	38.80	3649
Afro-Asiatic	0.3635	0.001643	95.75	358	36.52	5840
Algic	0.3826	0.002750	89.44	44	36.37	5577
Arawa	0.0389	0.000095	73.29	6	36.14	1882
Arawak	0.3527	0.002316	91.92	74	35.64	2437
Atlantic-Congo	0.4210	0.002034	95.91	1426	34.17	6037
Austroasiatic	0.3556	0.003215	87.77	170	37.67	3832
Austronesian	0.3965	0.003406	87.24	1256	35.12	3698
Barbacoan	0.0745	0.002130	79.66	6	39.00	3080
Border	0.2672	0.004609	85.08	15	33.00	3431
Bosavi	0.2519	0.003373	77.97	7	36.08	2176
Caddoan	0.3274	0.002654	87.26	5	36.57	4540
Carib	0.4274	0.003938	78.23	42	35.28	2301

Family	δ	*Q*-r	LDND	*N*	att	age
Central Sudanic	0.2242	0.002970	85.63	64	34.10	4618
Chibchan	0.4089	0.001915	93.25	27	36.70	4276
Chocoan	0.4580	0.036512	48.00	9	35.33	2175
Chukotko-Kamchatkan	0.0290	0.000057	70.35	5	39.40	3308
Cochimi-Yuman	0.4221	0.017798	75.72	9	31.42	2140
Daju	0.1327	0.001792	66.11	7	37.00	1811
Dogon	0.2685	0.005949	78.81	17	34.10	2235
Dravidian	0.3225	0.005260	77.26	81	34.23	2426
East Strickland	0.4988	0.018708	63.27	6	36.43	1398
Eastern Trans-Fly	0.7208	0.000569	87.78	4	36.92	3072
Eleman	0.4852	0.031169	65.55	6	37.71	5448
Eskimo-Aleut	0.2820	0.005962	75.69	11	37.60	5052
Eyak-Athapaskan-Tlingit	0.3397	0.006016	79.73	45	35.30	7483
Great Andamanese	0.2043	0.006331	74.95	10	36.63	2122
Guahibo	0.3294	0.003712	60.61	5	37.83	1291
Guaicuruan	0.2185	0.001378	86.39	5	37.60	2909
Gunwinyguan	0.4806	0.006485	90.25	9	32.67	3499
Heiban	0.3647	0.009270	73.34	10	39.18	2588
Hmong-Mien	0.3161	0.006042	86.38	37	36.88	3427
Huitotoan	0.0314	0.000187	68.92	5	37.20	2778
Ijoid	0.2817	0.007132	61.28	10	36.35	2639
Indo-European	0.2708	0.002194	89.31	479	35.75	4241
Iroquoian	0.4002	0.007503	79.62	11	36.71	4855
Japanese	0.0134	0.000032	61.28	13	38.67	1816
Je-Kaingang	0.3208	0.008893	79.98	21	34.77	5463
Jivaro	0.9922	0.009324	41.06	4	37.67	678
Kadugli-Krongo	0.4228	0.008419	56.28	6	34.00	1132
Kartvelian	0.0828	0.001957	78.95	5	39.25	2999
Khoe-Kwadi	0.3174	0.012814	72.94	13	35.88	3123
Kiwaian	0.3603	0.011184	61.06	6	37.50	1290
Koiarian	0.3511	0.001483	76.55	7	31.14	2640
Koman	0.1618	0.002754	89.15	4	37.00	3253

Family	δ	*Q*-r	LDND	*N*	att	age
Kwerba	0.4968	0.010236	90.07	8	29.40	3389
Lakes Plain	0.3585	0.003684	91.91	20	35.62	5290
Left May	0.3062	0.007723	80.95	6	30.86	2397
Lower Sepik-Ramu	0.3719	0.002244	96.23	32	31.00	6087
Maban	0.1005	0.001827	76.04	10	38.50	2382
Mande	0.2729	0.004468	83.01	72	35.19	3503
Mangarrayi-Maran	0.7595	0.015719	88.78	4	35.75	3725
Marind	0.1323	0.002490	88.29	6	30.92	3493
Matacoan	0.1785	0.003616	82.98	7	37.29	2569
Mayan	0.2420	0.005362	70.51	69	38.62	2186
Mirndi	0.3769	0.005430	87.58	5	36.75	3623
Miwok-Costanoan	0.2439	0.008668	70.98	10	36.78	3429
Mixe-Zoque	0.2524	0.007532	55.92	20	38.46	1432
Mongolic	0.3983	0.007010	77.25	16	35.00	2288
Muskogean	0.2184	0.007098	62.40	7	38.83	1720
Nadahup	0.1815	0.012661	67.96	4	34.00	1605
Nakh-Dagestanian	0.2483	0.002206	88.48	30	39.03	3969
Narrow Talodi	0.4898	0.011814	62.55	8	39.20	1402
Ndu	0.4535	0.013824	61.20	13	36.67	1314
Nilotic	0.2047	0.002665	86.73	51	37.08	4226
North Halmahera	0.4897	0.009483	69.58	16	34.61	1939
North Omotic	0.2688	0.004596	78.67	22	38.56	3129
Nubian	0.1203	0.001911	87.00	13	36.60	3500
Otomanguean	0.2651	0.001968	93.06	178	37.29	6580
Pama-Nyungan	0.3921	0.002725	92.89	223	33.94	4192
Panoan	0.4000	0.008551	71.92	38	36.58	1968
Pauwasi	0.1051	0.001568	88.97	5	35.14	4224
Pomoan	0.3951	0.025236	49.15	7	31.43	1216
Quechuan	0.2677	0.011004	39.94	45	38.05	839
Saharan	0.0191	0.000332	80.36	10	39.40	3938
Salishan	0.3019	0.003100	88.15	27	38.71	3713
Sepik	0.3374	0.002843	90.26	34	34.20	3901

Family	δ	*Q*-r	LDND	*N*	att	age
Sino-Tibetan	0.4102	0.002376	92.54	450	37.03	5376
Siouan	0.2231	0.002896	82.74	18	36.35	6187
Sko	0.3389	0.005387	90.17	9	33.75	4306
Songhay	0.3417	0.011538	50.31	10	39.25	1355
South Bird's Head Proper	0.0519	0.000804	71.21	6	34.38	1820
Surmic	0.2180	0.003292	82.17	10	38.11	3735
Tacanan	0.8984	0.011547	62.49	7	37.40	1590
Tai-Kadai	0.2887	0.005999	81.11	93	35.04	3519
Torricelli	0.3612	0.001884	95.96	53	35.35	5990
Totonacan	0.1337	0.004163	48.68	12	39.50	1365
Trans New Guinea	0.3942	0.001649	96.97	318	35.12	6883
Tucanoan	0.3029	0.008903	72.14	27	36.75	2694
Tungusic	0.3685	0.014542	59.59	12	35.35	1326
Tupi	0.3442	0.005248	79.83	73	35.17	3551
Turkic	0.3922	0.012833	61.42	43	36.19	1420
Uralic	0.2788	0.002564	85.54	40	38.46	3254
Uto-Aztecan	0.2404	0.003278	82.51	62	34.24	3802
West Bird's Head	0.4425	0.007261	82.74	5	36.11	2547
West Timor-Alor-Pantar	0.3853	0.004179	85.45	18	33.43	3685
Western Daly	0.1934	0.012806	48.84	11	28.78	1739
Yanomam	0.2362	0.004218	52.97	5	38.22	1021
Yeniseian	0.2448	0.005982	78.60	6	31.86	2693
Zaparoan	0.1884	0.002963	83.29	6	34.00	2596

Table 3. Spearman rank correlations for δ, *Q*-r, and family age across families.

Entities correlated	ϱ	*p*
δ ~ age	.0073	.9422
Q-r ~ age	-.4310	< .00001

3. Possible Correlates of Reticulation

In the following subsections we investigate whether each of the following properties influences reticulation, as measured by δ: age of families, status of a language as a linguistic isolate within a family, and the participation of a language in a dialect chain or a group of emerging languages.

3.1. *δ and Family Age*

In 2.3 above we found a negative correlation between *Q*-r and LDND, but no correlation between δ and LDND. The LDND score was measured as an average across all language pairs in each language group. Average LDND partly relates to the age of a family, but also to other factors. If a language group is 'explosive' in the sense that it contains many recently emerged languages, then the average LDND will tend to be low compared to a more 'implosive' group, containing many languages without particularly close relatives. So average LDND does not translate directly into age, and we will therefore take a more direct look at the latter. The hypothesis to be tested is the possibility mentioned by Gray et al. (2010) that families may become more treelike with age as extinction eliminates ambiguous cases. Age is here determined by the method of Holman et al. (2011): mean LDND is calculated between the highest-order subgroups of a given family and turned into an absolute age estimate using a calibrated logarithmic transformation. In Table 3 we show how δ and *Q*-r correlate with age across the 100 language families listed in Table 2. Not surprisingly, the influence of LDND on *Q*-r is strong enough to permeate the relationship between *Q*-r and age estimates. The complete absence of a correlation between δ and age is heartening, since it shows δ to not be an artifact of this particular characteristic of a language group (since age and att are uncorrelated, $\varrho = -.1414$, $p = .1607$, we do not need to control for the effects of the latter).

3.2. *δ and Phylogenetic Isolation*

The idea that treelikeness can arise from language extinction is perhaps attractive enough to warrant further testing. This hypothesis predicts that languages that are isolated within their family, seemingly having descended directly from the root, should be less reticulate. We now test this hypothesis for all languages in the ASJP database (Version 14) that are within-family isolates according to *Ethnologue*. Languages designated as 'unclassified' within their families are not included in our sample of isolates because their position is uncertain. We use *Ethnologue* for the internal classification of families since Hammarström's classification is not fully developed in this respect and largely follows *Ethnologue*. Many of *Ethnologue*'s within-family isolates are probably real isolates without any close relatives whatsoever (e.g., Hatsa and Sandawe, which are both considered to be Khoisan in *Ethnologue*), and these are inevitably going to show high reticulation. So for family definitions we use the more conservative Hammarström classification. In Table 4 we provide δ for each within-family isolate as the average δ of all quartets in which the language participates, so that this score can be compared to the average delta for all quartets in the family, provided in Table 2. We exclude families where more than half of the members are not further subclassified. We also exclude families where fewer than four ISO 639-3 languages other than isolates are represented in the database, because these families lack quartets without isolates for comparison. As already mentioned in the introduction to this paper, when the ASJP database contains more than one word list for an ISO 639-3 language, one list is arbitrarily chosen for the present study. Word lists in the ASJP database are uniquely identified by ASJP names of languages. In Table 4 we indicate by means of stars preceding the names in the column 'ASJP name' those cases where the language name (= word list) was arbitrarily selected among different representatives of a given ISO 639-3 language. Table 4 also provides a column with the number of attested items in the word lists for each of the isolates.

Table 4. Data on reticulation for isolates within families represented in the ASJP database.

Ethnologue name	ISO 639-3	ASJP name	Family	δ for language	att for language
Ubykh	uby	Ubykh	Abkhaz-Adyge	0.0327	35
Wiyot	wiy	Wiyot	Algic	0.3486	33
Yurok	yur	Yurok	Algic	0.3615	40
Tlingit	tli	Tlingit	Athapaskan-Eyak-Tlingit	0.3270	37
Bunun	bnn	Bunun	Austronesian	0.4076	34
Paiwan	pwn	Paiwan	Austronesian	0.4246	34
Puyuma	pyu	*Nanwang Puyuma	Austronesian	0.4175	32
Rukai	dru	*Mantauran	Austronesian	0.4337	31
Caddo	cad	Caddo	Caddoan	0.3898	40
Barí	mot	Bari Columbia	Chibchan	0.4579	35
Pech	pay	Pech	Chibchan	0.4409	32

Ethnologue name	ISO 639-3	ASJP name	Family	δ for language	att for language
Woun Meu	noa	Wounaan	Choco	0.5182	39
Cochimi	coj	Cochimi	Cochimi-Yuman	0.4412	30
Kiliwa	klb	Kiliwa	Cochimi-Yuman	0.3584	38
Paipai	ppi	Paipai	Cochimi-Yuman	0.4635	29
Havasupai-Walapai-Yavapai	yuf	*Havasupai	Cochimi-Yuman	0.4759	29
Aleut	ale	Aleut	Eskimo-Aleut	0.3543	39
Djauan	djn	Djauan	Gunwinyguan	0.4847	35
Buan	ngk	Buan	Gunwinyguan	0.5107	36
She	shx	*Chenhu She	Hmong-Mien	0.3191	39
Defaka	afn	Defaka	Ijoid	0.2400	33
Armenian	hye	*Eastern Armenian	Indo-European	0.3399	40
Cherokee	chr	Cherokee	Iroquoian	0.3214	29
Awera	awr	Awera	Lakes Plain	0.3690	38
Kambot	kbx	*Kambot/ Kambaramba	Lower Sepik-Ramu	0.4613	28
Mogholi	mhj	Moghol	Mongolic	0.4049	34
Dargwa	dar	Dargwa	Nakh-Daghestanian	0.3113	38
Khinalugh	kjj	Khinalug	Nakh-Daghestanian	0.2791	39
Lak	lbe	Lak	Nakh-Daghestanian	0.2996	39
West Makian	mqs	*Bobawa	North Halmahera	0.4400	32
Nobiin	fia	Nobiin	Nubian	0.1167	33
Midob	mei	Midob	Nubian	0.3495	39
Chiapanec	cip	Chiapanec	Otomanguean	0.3219	34
Bandjalang	bdy	*Gidabal	Pama-Nyungan	0.3928	38
Kumbainggar	kgs	Gumbaynggir	Pama-Nyungan	0.4256	40
Kala Lagaw Ya	mwp	Kala Laggaw Ya	Pama-Nyungan	0.4165	35
Muruwari	zmu	Muruwari	Pama-Nyungan	0.4093	35
Yanyuwa	jao	Yanyuwa	Pama-Nyungan	0.4202	40
Yugambal	yub	Yugambal	Pama-Nyungan	0.3872	38

Ethnologue name	ISO 639-3	ASJP name	Family	δ for language	att for language
Kaxararí	ktx	Kaxarari	Panoan	0.4022	39
Cashibo-Cacataibo	cbr	Cashibo	Panoan	0.4063	32
Southeastern Pomo	pom	Southeastern Pomo	Pomoan	0.4403	28
Bella Coola	blc	Bella Coola	Salishan	0.3150	38
Tillamook	til	Tillamook	Salishan	0.3617	37
Catawba	chc	Catawba	Siouan	0.2576	36
Majang	mpe	Mesengo	Surmic	0.1904	39
Wiru	wiu	Wiru	Trans-New Guinea	0.4107	40
Cubeo	cub	*Cubeo	Tucanoan	0.2721	40
Awetí	awe	Aweti	Tupi	0.3866	28
Sateré-Mawé	mav	Satere Mawe	Tupi	0.3240	37
Chuvash	chv	Chuvash	Turkic	0.4249	40
Hungarian	hun	*Csango	Uralic	0.3256	35
Khanty	kca	Khanty	Uralic	0.2991	40
Mansi	mns	Mansi	Uralic	0.2870	40
Bunak	bfn	Bunak	West Timor-Alor-Pantar	0.4113	31
Wersing	kvw	Wersing	West Timor-Alor-Pantar	0.3833	32
AVERAGE				0.3674	

Within-family isolates are more reticulate than the family average in the majority of the cases listed in Table 4, i.e., in 43 out of a total of 56. A paired t-test across the 56 cases shows that the tendency for within-family isolates to be more reticulate than the family averages is significant (t = 3.5879, p = .0007). (For this test there is no need to control for att, since the difference in att for isolates and for the families to which they belong is uncorrelated with the difference between δ for isolates and the families to which they belong; ϱ = -.0408, p = .7654).

Thus, we can reject the hypothesis that within-family isolates, which are likely to be remnants of erstwhile larger groups whose other members became extinct, behave in a more treelike fashion. In fact, the opposite picture appears. Presumably a common way for an isolate to arise is through the extinction of its closest relatives. With the loss of all but one member of a linguistic subgroup, information is also lost about the phylogenetic paths previously linking the language to the larger family network. In the light of these considerations it is not unexpected that within-family isolates should be more reticulate.

3.3. *δ and Dialects or Emerging Languages*

A valid question, complementary to the one posed in the previous section, is whether groups of dialects or emergent languages are more reticulate than languages which are less closely related. We test this in two ways. First, we test whether dialects tend to be more or less reticulate than the common language they represent; next, we test whether groups of emerging languages—i.e., groups of speech varieties that are somewhat more differentiated than dialects—tend to be more or less reticulate than the average of the families to which they belong. Intuitively, if languages are still in the process of splitting up, they may be undergoing mutual borrowing to a larger extent than languages which have been mutually unintelligible for hundreds or thousands of years, and are therefore perhaps expected to be more reticulate. On the other hand, if Atkinson et al. (2008) are correct in claiming that the speed of language change increases when languages split up, then we would perhaps expect emerging languages to be less reticulate. According to the model of Atkinson et al., speakers will enhance the aspects of their speech that are different from the speech of closely related varieties rather than seek to keep their speech similar to that of the neighbors.

For the investigation of whether dialects tend to be more or less reticulate than the language they represent, we sample all the groups of word lists that pertain to the same ISO 639-3 code according to *Ethnologue*. Although there is of course no commonly accepted definition of a dialect as opposed to a language, we cannot err completely by regarding speech forms pertaining to the same ISO 639-3 code as being dialects of the same language. *Ethnologue* tends to regard many speech forms considered to be dialects by experts as being distinct languages. So we are not likely to be sampling speech forms that linguists would consider distinct languages when sampling groups of speech forms pertaining to the same ISO 639-3 code. For each such group having 4 or more members we calculate the average δ within the group and compare it to δ for the language as a whole within its family. δ for the 'language as a whole' is, as in the previous section, the average δ for all the quartets within the Hammarström family to which an arbitrary representative of the ISO 639-3 code pertains. Results are shown in Table 5. This provides the ISO-code ('ISO'), the name of the arbitrary representative in the ASJP database ('name'), the number of members of the ISO-code group ('*N*'), δ and average number of attested items for the ISO-code group ('δ ISO' and 'att ISO'), and δ and number of attested items for the arbitrary representative within the family ('δ rep' and 'att rep'). The table is ordered alphabetically by ISO-codes.

Table 5. Data on reticulation for dialects as opposed to languages.

ISO	Name	*N*	δ ISO	att ISO	δ rep	att rep
abl	Abung Sukadana Lampung Nyo	4	0.4659	32.25	0.4153	33
abz	Abui/Atimelang	4	0.7542	34.50	0.3903	32
anv	Denya/Bajwo	4	0.1145	34.00	0.4472	30
apb	Aulu Saa	4	0.2987	33.00	0.3988	33
auw	Awyi Unknown Dial	5	0.0839	33.80	0.2219	37
baa	Avaso Babatana	5	0.4130	33.00	0.4088	33
bao	Bara	4	0.5127	35.50	0.2956	40
bca	Dashi Bai	11	0.3437	38.09	0.4024	39
bdl	Anaiwoi Bajau	17	0.4078	32.06	0.3847	32
beu	Blagar/Apuri Pura	6	0.1905	31.83	0.4018	32
bfa	Bari Sudan	4	0.0954	40.00	0.2056	40
bfc	Ega Bai	9	0.3238	38.33	0.4212	39
bfs	Hedian Bai	8	0.3360	38.25	0.4002	36
bft	Chorbat Balti	9	0.3233	30.67	0.4102	28
bhk	Buhi	5	0.5688	38.60	0.3385	40
bjq	Malagasy Antaisaka	9	0.2978	37.33	0.4123	39
bmg	Libobi	4	0.5707	33.00	0.4018	34
bnk	Bierebo Bonkovia	4	0.7141	36.25	0.4242	37
bod	Lhasa Tibetan	4	0.9899	39.75	0.3733	40
bon	Bine/Boze Giringarede	16	0.3366	37.63	0.7208	39
bqz	Babong	5	0.5992	40.00	0.4253	40
btr	Baetora Narovorovo	5	0.5672	35.60	0.3794	35
bvx	Botongo Dibole	5	0.2856	32.80	0.4082	32
bzp	Arandai/Barau	4	0.0404	35.00	0.0519	37
cdr	Cinda 1	5	0.1191	32.40	0.4372	33
cia	Batu Atas	6	0.4438	39.33	0.3961	39
elp	Elpaputih Samasuru Paulohij	4	0.0203	36.75	0.4099	35
evn	Evenki Poligus Literary	5	0.5313	33.60	0.3544	40

ISO	Name	*N*	δ ISO	att ISO	δ rep	att rep
gdr	Abam	15	0.4115	36.27	0.7208	38
ggo	Adilabad Gondi	5	0.3041	34.40	0.2614	34
gio	Gelao	5	0.3128	33.40	0.2567	37
gju	Agra Gujari	13	0.3918	33.38	0.2386	34
gri	Ghari	6	0.4381	32.67	0.3926	34
gup	Gunwinggu Manyallaluk Mayali	5	0.5176	29.20	0.4926	28
gwd	Gawwada	4	0.5223	39.00	0.3479	40
gwn	Arabishi	6	0.4701	39.83	0.4144	39
hig	Bazza	7	0.3929	34.71	0.3623	35
hnd	Attock City Hindko	8	0.3239	34.25	0.2415	35
hno	Balakot Hindko	6	0.3133	33.33	0.2464	35
ijc	Apoi	24	0.3243	36.79	0.3783	37
ium	Chiangrai Mien	4	0.0754	38.00	0.2514	39
ivv	Imorod	6	0.2443	32.60	0.3875	31
jbj	Arandai/Najarago	5	0.0868	33.80	0.0576	34
jod	Nowolokakan	5	0.2087	32.00	0.2314	32
jpn	Japanese 2	5	0.4759	38.80	0.0134	40
kcf	Iigau	6	0.2515	34.33	0.4357	32
ken	Bas Kenyang	4	0.0103	36.25	0.4394	36
kga	Kanikakan	5	0.2045	32.00	0.2345	32
kge	Adumanis Ulu Komering	6	0.3234	32.83	0.4134	33
khb	Jinghong Tai Lue	5	0.3033	38.60	0.3004	40
khw	Chatorkhand Khowar	6	0.3129	34.50	0.3410	34
kiw	Anigibi	4	0.1564	37.75	0.5142	38
kjd	Domori	5	0.3707	37.40	0.4116	34
klz	Kabola	4	0.6860	33.75	0.3312	39
kmz	Asadli Khorasani	22	0.4152	35.68	0.3716	36
kvg	Boazi/Boazi	6	0.3823	29.83	0.1323	33
lbw	Tolaki	6	0.5170	38.33	0.3986	30
led	Lendu	6	0.3050	29.83	0.2193	35
lev	Lamma	4	0.0517	32.75	0.4218	37
ljp	Belalau Lampung Api	14	0.3708	32.07	0.4247	31

ISO	Name	*N*	δ ISO	att ISO	δ rep	att rep
llp	North Efate Nguna	5	0.2130	36.80	0.4290	37
lww	Lewo Filakara	7	0.2631	36.29	0.4088	37
lyn	Kwandi	8	0.3403	34.13	0.3549	32
mam	Mam Northern	10	0.4573	39.40	0.2505	40
mda	Gbugyar	9	0.3939	33.78	0.4330	32
mgd	Agi	7	0.2665	36.71	0.1843	37
mky	East Makian	9	0.2089	32.56	0.4213	37
mms	Mam Cabrican	7	0.4117	39.29	0.2261	39
mqs	Bobawa	8	0.2750	32.88	0.4400	32
mww	Hmong Daw	4	0.8902	35.50	0.2633	34
mxx	Baralakakan	6	0.2568	32.00	0.2635	32
ngc	Doko	4	0.0742	33.00	0.4302	34
ngu	Nahuatl Acatlan	5	0.2765	30.60	0.2643	29
nij	Kapuas Kahayan	5	0.3201	35.60	0.3702	38
nmk	Namakura Bongabonga	4	0.6446	36.50	0.4044	38
nwi	Southwest Tanna Enfitana	5	0.4284	34.40	0.4082	35
pbt	Chaman Pashto	5	0.6378	35.00	0.2816	35
pbu	Baffa Pashto	25	0.3468	35.08	0.3117	35
pcc	Po Ai	4	0.8920	35.25	0.3125	31
plt	Malagasy Ambositra	7	0.2663	37.71	0.4044	38
pnp	Kamboa	5	0.2397	39.60	0.4119	40
pst	Bannu Pashto	5	0.4404	34.60	0.2983	35
quc	Central Quiche	7	0.3220	37.00	0.2356	40
qug	Chimborazo Quichua	7	0.3301	39.29	0.3061	38
qut	Kichee Aldea Argueta Solola	16	0.3448	37.75	0.2291	38
qvi	Quechua Imbabura	4	0.3369	39.50	0.2580	40
rmc	Burgenland Romani	6	0.4739	38.83	0.2462	40
rmn	Bugurdzi Romani	7	0.4149	39.00	0.2602	40
rmy	Banatiski Gurbet Romani	6	0.5166	39.83	0.2755	39
sat	Bodobelghoria Santali	8	0.4882	36.50	0.3265	37
shb	Ninam	4	0.2348	37.75	0.2362	34
shx	Chenhu She	4	0.0382	34.75	0.3191	39

ISO	Name	*N*	δ ISO	att ISO	δ rep	att rep
sie	Liuwa	9	0.3218	33.44	0.3511	34
skg	Malagasy Sakalava 1	8	0.2358	37.00	0.4199	38
ssw	Hlubi	4	0.1047	37.00	0.4071	40
stc	Banua	5	0.4999	34.60	0.4010	37
str	Saanich	4	0.0798	38.50	0.2847	39
swh	Swahili Chirazi	8	0.2492	32.13	0.3543	31
swi	Sandong Sui	5	0.2313	33.83	0.2709	39
tad	Deirate	5	0.1924	33.40	0.4328	31
tay	Atayal	4	0.0332	35.25	0.4168	40
tcc	Datooga Dialect 2	5	0.1322	31.00	0.2116	31
tdx	Malagasy Mahafaly	5	0.3581	37.80	0.4189	39
tha	Siamese	6	0.4353	36.33	0.2856	38
tlr	Koo Talise	5	0.2665	33.60	0.4050	30
tof	Gizra/Kupere	6	0.2977	36.83	0.7208	36
twe	Teiwa	5	0.4842	33.40	0.4070	39
tzm	Figuig	4	0.4038	38.75	0.3285	38
upv	Atchin	7	0.2810	35.14	0.3718	36
woi	Kamang	6	0.4854	31.83	0.3965	32
wsi	Wusi Kerepua	4	0.2799	35.00	0.4030	34
xbr	Kambera	4	0.1499	34.00	0.4237	34
xho	Mpondo	5	0.4229	38.40	0.4074	40
xmv	Malagasy Antankarana	4	0.9149	36.50	0.4074	38
zmx	Bene Bomitaba	16	0.3245	34.81	0.4067	35
zyb	Tai Wuming	4	0.0026	34.75	0.2981	34
zzj	Lung Chow	8	0.3771	35.13	0.2831	30

Among the 117 cases in Table 5, there are 58 cases where the dialect group has a greater mean δ than the language which the dialects represent, and 59 cases where the opposite situation holds. This distribution is maximally balanced, so hypotheses of either more or less reticulation in dialects are both rejected by a paired t-test ($t = -.1592$, $p = .8738$). (Before carrying out this test we checked whether attestations could be a confounding factor, but this was not the case, since the difference between δ ISO and δ rep is uncorrelated with the difference between att ISO and att rep; $\varrho = -.1020$, $p = .2738$).

We now supplement this analysis with one looking at whether emerging languages, i.e., speech forms which are closely related but less similar than dialects, are more or less reticulate than the average for languages within the families to which the emerging languages pertain. In order to draw an appropriate sample uncontaminated by our own biases, we assume that languages grouped in the category 'macrolanguage' in *Ethnologue* (Lewis, 2009) are examples of emerging languages. This assumption is based on the online *Ethnologue*'s definition of macrolanguages as "multiple, closely related individual languages that are deemed in some usage contexts to be a single language."

In Table 6 we present data for all ASJP word lists representing languages that are classified as belonging to macrolanguages. As usual, only one word list is used per ISO 639-3 code, and the family classification is Hammarström's. The first column names the macrolanguage; the second indicates languages represented in the ASJP database (but, for economy of space, gives no details about the exact word lists used); '*n*' indicates the number of macrolanguage members in the database; '*N*' indicates the total number of languages in each macrolanguage group. The last four columns give averages of δ and attestations for the macrolanguage representatives and for the family as a whole, respectively.

Table 6. Data on reticulation for members of macrolanguages.

Macrolanguage	Representatives in database (ISO)	*n*	*N*	δ macro-language	att macro-language	δ family	att family
Albanian	als	1	4	0.3268	40.00	0.2708	35.75
Arabic	arq, shu, acy, arz, afb, ayl, acm, ary, apc, ayn, ajp, apd, aeb	13	30	0.3156	36.85	0.3635	36.52
Azerbaijani	azj, azb	2	2	0.3926	34.00	0.3922	36.19
Baluchi	bgp, bgn	2	3	0.3015	33.50	0.2708	35.75
Bikol	bhk, bcl, bto, cts, bln	5	5	0.3460	35.80	0.3965	35.12
Buriat	bxm	1	3	0.4875	30.00	0.3983	35.00
Chinese	hak, cmn, nan, wuu, hsn, yue	6	13	0.4044	39.33	0.4102	37.03
Delaware	umu, unm	2	2	0.3744	37.50	0.3826	36.37
Dinka	dib, dks, dik	3	5	0.2203	35.67	0.2047	37.08
Fulah	fub, ffm, fuv, fuc, fuf	5	9	0.4354	35.80	0.4210	34.17
Gbaya	gya, gso, gbp	3	6	0.4386	31.00	0.4210	34.17
Gondi	gno, ggo	2	2	0.2695	33.50	0.3225	34.23

Macrolanguage	Representatives in database (ISO)	*n*	*N*	δ macro-language	att macro-language	δ family	att family
Guaraní	gui, gun, gug	3	5	0.3579	38.00	0.3442	35.17
Hmong	hnj, hmm, cqd, hea, mmr, mww	6	24	0.3310	36.17	0.3161	36.88
Inuktitut	ike, ikt	2	2	0.2126	37.50	0.2820	37.60
Inupiaq	esi	1	2	0.2215	38.00	0.2820	37.60
Kalenjin	sgc, niq, oki, pko, spy, tuy	6	9	0.2032	36.50	0.2047	37.08
Kanuri	knc, kby	2	3	0.0165	39.50	0.0191	39.40
Komi	koi, kpv	2	2	0.2614	36.00	0.2788	38.46
Kongo	kng, ldi	2	3	0.3959	35.50	0.4210	34.17
Konkani	knn	1	2	0.2884	28.00	0.2708	35.37
Kpelle	gkp, xpe	2	2	0.2725	39.00	0.2729	35.19
Kurdish	ckb, kmr	2	3	0.2919	38.50	0.2708	35.75
Lahnda	hno, hnd, pnb, skr	4	8	0.2452	34.50	0.2708	35.75
Malagasy	xmv, bhr, msh, bmm, plt, skg, bjq, tdx, txy, xmw	10	10	0.4118	36.90	0.3965	35.12
Malay	zlm, xmm, max	3	15	0.3668	33.67	0.3965	35.12
Mandingo	mnk, myq, mlq	3	7	0.2529	37.33	0.2729	35.19
Mari	mrj, mhr	2	2	0.2936	37.50	0.2788	38.46
Mongolian	khk	1	2	0.3267	39.00	0.3983	35.00
Ojibwa	ciw, ojg, ojs	3	7	0.3479	39.33	0.3826	36.37
Oromo	orc, gax, hae, gaz	4	4	0.3259	39.00	0.3635	36.52
Pushto	pst, pbu, pbt	3	3	0.2972	35.00	0.2708	35.75
Rajasthani	gju	1	6	0.2386	34.00	0.2708	35.75
Romany	rmn, rml, rmc, rmf, rmo, rmy, rmw	7	7	0.2550	39.57	0.2708	35.75
Serbo-Croatian	bos, hrv, srp	3	3	0.2460	40.00	0.2708	35.75
Slave	scs	1	2	0.3382	35.00	0.3397	35.30
Swahili	swh	1	2	0.3543	31.00	0.4210	34.17
Syriac	cld	1	2	0.3333	40.00	0.3635	36.52
Tamashek	thv, ttq, thz, taq	4	4	0.3360	38.25	0.3635	36.52
Uzbek	uzn	1	2	0.4328	40.00	0.3922	36.19

Macrolanguage	Representatives in database (ISO)	*n*	*N*	δ macro-language	att macro-language	δ family	att family
Yiddish	add	1	2	0.2265	37.00	0.2708	35.75
Zapotec	zaq, zpo, zaf, zad, zpc, zai, zpl, ztp, zaw, zpm, zac, ztq, zpx, zab, zpf, zpn, zpi, zaa, zpz, zts, ztg, zpu, zae, zav, zpq	25	57	0.2510	39.04	0.2651	37.29
Zaza	diq, kiu	2	2	0.2919	36.00	0.2708	35.75
Zhuang	zch, zgn, zyb, zzj	4	16	0.2917	36.25	0.2887	35.04
AVERAGE				0.3097	36.45	0.3190	35.98

We now determine whether members of macrolanguages tend to be significantly more or less reticulate than their family averages, again using a paired t-test. The result is that there is no significant trend in either direction (t = -1.7879, p = .0808). Since the number of languages in the database (n) is in many cases small, we also tested whether significance could be reached when requiring that n should exceed a certain cut-off point for the data point to be included. It turned out, however, that there is no value of n for which $p < .05$. (And, as before, we tested for att as a confounding factor by correlating the difference between delta for macrolanguage and for family with the difference between att for macrolanguage and att for family, not finding a significant correlation: ϱ = -.1222, p = .4294). Thus, we conclude that dialects and emerging languages are neither more nor less reticulate than languages at large.

4. Consequences of Reticulation for Classification

In the beginning of this paper we found δ (Holland et al., 2002) to be a promising measure of reticulation, while *Q*-r (Gray et al., 2010) looked less promising. In the absence of tests on a wide range of phylogenetic cases across different disciplines, it is possible that δ is more sensitive to artificial effects of distributional aspects of distances in the matrix than to real-world factors inducing conflicts in the phylogenetic signal. Our findings in Sections 2.3 and 3.1, however, strongly suggest that this is not the case: unlike *Q*-r, δ is not sensitive to the heterogeneity or age of language families, and it is not sensitive to the size of families either. Sensitivity to such factors would indicate a weakness in the measure, since there is no obvious reason why they should influence reticulation.

Our subsequent finding that languages which are isolated within their respective phylogenies tend to be more reticulate could potentially be an artifact somehow following from the distribution of linguistic distance values. However, we would then expect the opposite situation, that of dialects or recently emerged languages, to correspond to a smaller amount of reticulation; but we did not find any significant trend in δ values in this situation. These findings alone suggest that δ does seem to be measuring real effects rooted in events involving the social histories of speakers.

There is another completely different line of evidence showing that δ measures real socio-historical processes related to reticulation, while *Q*-r does not. It is known that classifications of different language families based on distance measures, including the version of the Levenshtein distance (LDND) currently used by ASJP, vary with respect to the degree to which they conform to the classifications of experts (Wichmann et al., 2010a; Huff and Lonsdale, 2011; Pompei et al., 2011; Greenhill, 2011). Here we are going to present evidence that the performance of ASJP, as measured by the conformity of ASJP classifications for different language families with those of experts, correlates with δ and not with *Q*-r. The issue here is not to decide which classification is best for each family, something far beyond the scope of this paper. Rather, what we are going to suggest is that differences in classifications should be interpreted as controversies arising from real conflicts in phylogenetic signals, which ultimately have real socio-historical causes.

The best published evaluation of the performance of ASJP classifications to date is that of Pompei et al. (2011). Wichmann et al. (2010a) and Huff and Lonsdale (2011) use less sophisticated measures of differences between trees. The results of Greenhill (2011) are less useful for present purposes because they are difficult to replicate and evaluate from the available documentation of data and methods, and also because that study is concerned with Austronesian languages only.

Like Wichmann et al. (2010a) and Huff and Lonsdale (2011), Pompei et al. compare trees derived from ASJP distance measures with the *Ethnologue* classification, but, as a new feature, they modify the measure of differences between trees to take into account tree resolution. Two measures of distances among trees are used. One is called Generalized Robinson-Foulds (GRF). It counts the nodes in one tree that are not in the other tree and vice versa, but a difference is not counted when a node in one tree introduces a split between taxa whose relationships remain unresolved in the other tree. This takes into account the fact that trees based on distance data like those of ASJP will be almost fully binary—a quality for which they should not be punished—while the expert trees of linguistics are often not very resolved. Another measure is the Generalized Quartet Distance (GQD). This counts the number of 'butterflies' not shared by two trees divided by the number of butterflies in the expert tree, in this case that of *Ethnologue*. A 'butterfly' is defined as a resolved quartet—one that is not starlike (Christiansen et al., 2006). The GQD, again, avoids punishing a tree to be tested for being more resolved than the tree it is tested against (for more discussion and further refinement of the GQD cf. Walker et al., 2012).

Thus, in the following our point of departure is the results of Pompei et al. (2011), but we additionally use two alternative performance measures introduced in Holman et al. (2008) and used again in Wichmann et al. (2010a). These compare ASJP distances directly to distances derived from the topology of *WALS* and *Ethnologue* trees, rather than making indirect comparisons based on ASJP trees that have passed through the filter of a phylogenetic algorithm. The measure of similarity used for comparing ASJP distances to the *WALS* classification is r, the standard Pearson product-moment correlation (across pairs of languages), where taxonomic distance is defined as 1 for languages in the same genus, and 2 for languages in different genera but the same family. The measure of similarity used for comparing ASJP distances to *Ethnologue* is the Goodman-Kruskal γ, which is defined as $(C-D)/(C+D)$, where C is the number of concordant comparisons (those ordered in the same direction on both variables), and D is the number of discordant comparisons (those ordered in opposite directions on the two variables). In the present application, one variable is ASJP distance and the other is taxonomic distance in *Ethnologue*. For the latter, if *Ethnologue* classifies two languages in the same group and a third language outside that group, then the distance between the first two languages is less than the distance between the first and third languages, and also less than the distance between the second and third languages. γ summarizes the consistency of such distances with ASJP distances. Like other correlation coefficients, γ ranges from -1 to +1, and takes the value 0 if the variables are independent. The fact that γ is usually higher than the Pearson correlation merely reflects the fact that the former ignores ties, which are frequent in taxonomic distances. The reason why two different correlation measures are used is that the *WALS* classification operates with just two taxonomic levels within families (the language and the genus), which are intended to be comparable across families; in contrast, the *Ethnologue* classification is more complicated and involves variable numbers of taxonomic levels, which are not intended to be comparable across families or even across branches of the same family.

We first base analyses on the same dataset that Pompei et al. (2011) used, and then repeat them on the most recent published version of the ASJP database, also introducing the additional differences of using Hammarström's family definitions rather than family definitions from *WALS* (Haspelmath et al., 2005) and basing trees used for the tree comparison scores on Neighbor-Joining rather than the new algorithm, FastSBiX, which is described in Tria et al. (2010a, b) but not yet implemented in publicly available software.

First, then, in Table 7 we compare δ and *Q*-r, the two measures of reticulation, with GRF, GQD, *r*, and γ, the four different measures of comparison between ASJP results and expert classifications. Pompei et al. (2011) use the family definitions of *WALS* but the internal classification of *Ethnologue*, so we do the same here. Note that GRF and GQD are measures of mismatches, so they increase with a poorer fit, whereas *r* and γ increase with a better fit. Pompei et al. tested different methods of deriving a distance measure from string comparisons and also tested different phylogenetic algorithms. The GRF and GQD values cited here are those deriving from the best-performing distance measure, which is the standard LDND normally used within the ASJP project, and from the best-performing phylogenetic algorithm, FastSBiX. Pompei et al. drew upon data from Version 12 of the ASJP Database (Wichmann et al., 2010b), including the cases of more than one list for a single language, so δ, *Q*-r, *r*, and γ are also based on this dataset; the values for *r* and γ were already given in Wichmann et al. (2010a), but are repeated here for convenience. The number of word lists available for each language in Version 12 of the ASJP Database is provided both in Wichmann et al. (2010a) and Pompei et al. (2011) and is not repeated here. Only families for which 10 or more word lists were available are reported on here. For families with a single genus, *r* is not defined and the corresponding cell in the table is therefore blank. There are 49 families with at least 10 lists, of which 33 have at least two genera.

Table 7. Reticulation and fit between ASJP and expert classifications for *WALS* families (at least 10 members).

Family	δ	*Q*-r	GRF	GQD	*r*	γ	att
Afro-Asiatic	0.3594	0.001646	0.2149	0.0323	0.6444	0.7375	36.83
Algic	0.3739	0.003391	0.3462	0.3768	0.3969	0.5459	36.86
Altaic	0.2243	0.004940	0.3378	0.1102	0.8711	0.9240	35.48
Arawakan	0.3763	0.003445	0.1707	0.1407		0.4934	36.54
Australian	0.4199	0.002506	0.3653	0.2230	0.3020	0.4463	33.84
Austro-Asiatic	0.3430	0.003534	0.3265	0.0757	0.5942	0.6459	37.10
Austronesian	0.3951	0.002131	0.3723	0.1907	0.1589	0.2535	34.74
Border	0.2602	0.005043	0.0000	0.0000		0.7763	32.94
Bosavi	0.2214	0.004218	0.0000	0.0000		0.9369	35.87
Cariban	0.4091	0.004144	0.8750	0.5458		0.2879	36.89
Chibchan	0.4029	0.001971	0.5333	0.2488	0.5796	0.6935	36.25
Dravidian	0.3496	0.009410	0.3889	0.1715	0.3612	0.5246	33.52

Family	δ	*Q*-r	GRF	GQD	*r*	γ	att
Eleman	0.2251	0.009813	0.0000	0.0000	0.9304	0.9574	36.70
Great Andamanese	0.2051	0.007037	0.2857	0.6786		0.1974	36.80
Hmong-Mien	0.2346	0.006285	0.2727	0.1316		0.9333	36.43
Hokan	0.2673	0.006460	0.4000	0.1099	0.8539	0.5320	32.58
Indo-European	0.2346	0.002218	0.3084	0.0662	0.7529	0.8251	37.15
Kadugli	0.3902	0.013792	0.0000	0.0000		0.8039	34.00
Khoisan	0.2267	0.003211	0.4615	0.1923	0.7047	0.6899	35.07
Kiwaian	0.2957	0.011868	0.0000	0.0000		0.9441	37.60
Lakes Plain	0.3083	0.003655	0.1739	0.2096	0.4219	0.7100	35.62
Lower Sepik-Ramu	0.2675	0.002474	0.0000	0.0000	0.6054	0.9282	30.95
Macro-Ge	0.3064	0.002448	0.3810	0.2310	0.6887	0.6797	34.29
Marind	0.1561	0.006096	0.0690	0.0284	0.6345	0.9370	33.03
Mayan	0.2328	0.002572	0.1549	0.0327		0.8236	39.13
Mixe-Zoque	0.2309	0.001732	0.2222	0.0754		0.9803	38.36
Morehead & U. Maro Rivers	0.1986	0.005478	0.0714	0.1540		0.6930	33.00
Na-Dene	0.3196	0.001933	0.5789	0.2098	0.6387	0.7728	35.14
Nakh-Daghestanian	0.2489	0.002291	0.1034	0.0653	0.6621	0.9397	39.03
Niger-Congo	0.4302	0.006966	0.4127	0.1078	0.4335	0.4021	34.15
Nilo-Saharan	0.3363	0.001544	0.2752	0.0951	0.6209	0.5830	36.47
Otomanguean	0.2647	0.002045	0.0357	0.0015	0.8507	0.9906	37.90
Panoan	0.3599	0.007509	0.7333	0.4805		0.3802	36.67
Penutian	0.2751	0.002374	0.0556	0.0066	0.8760	0.8156	35.81
Quechuan	0.3143	0.008604	0.0000	0.0000		0.4565	39.56
Salishan	0.2548	0.004347	0.1111	0.0600	0.6521	0.8903	36.83
Sepik	0.2783	0.002664	0.0435	0.0468	0.6514	0.8618	34.54
Sino-Tibetan	0.3848	0.002710	0.4130	0.1012	0.5922	0.6942	37.78
Sko	0.1848	0.002124	0.0000	0.0000	0.8193	0.7450	32.93
Tai-Kadai	0.2978	0.009206	0.4118	0.2232	0.6840	0.7725	35.41
Torricelli	0.2516	0.001646	0.2500	0.1144	0.6037	0.8909	35.32
Totonacan	0.1948	0.002465	0.0000	0.0000		1.0000	39.43
Trans-New Guinea	0.4131	0.005278	0.2265	0.1078	0.5065	0.6748	34.83

Family	δ	*Q*-r	GRF	GQD	*r*	γ	att
Tucanoan	0.2886	0.008674	0.1875	0.1105		0.7565	39.63
Tupian	0.3409	0.005472	0.4667	0.1517	0.7594	0.9185	35.36
Uralic	0.2700	0.004796	0.1500	0.0373	0.5057	0.9742	39.00
Uto-Aztecan	0.2984	0.002490	0.1892	0.0976	0.9189	0.7566	32.83
West Papuan	0.2101	0.004213	0.1613	0.0550	0.6093	0.7432	35.06
Western Fly	0.1718	0.002746	0.0000	0.0000		1.0000	36.67

Table 8. Spearman correlations for reticulation (δ and *Q*-r) vs. fit with expert classifications (GRF, GQD, *r*, and γ) across *WALS* families.

Entities correlated	ϱ	*p*
GRF ~ δ	.5144	.0002
GQD ~ δ	.4220	.0025
r ~ δ	-.6273	.0001
γ ~ δ	-.5982	< .00001
GRF ~ *Q*-r	-.0649	.6579
GQD ~ *Q*-r	.0140	.9240
r ~ *Q*-r	-.0201	.9544
γ ~ *Q*-r	-.1206	.4093

Table 8 provides data on correlations between relevant columns of Table 7. It shows that δ, but not *Q*-r, correlates highly with all four measures of fit between expert classifications and ASJP classifications. (None of the columns shows a significant correlation with the average number of attestations, absolute values of ϱ ranging from .0187 to .2345, and $.1048 < p < .9177$, so this is not a confounding factor here).

We now repeat the exercise using the more recent expanded and corrected version of the database (Wichmann et al., 2011) and the conservative family definitions of Hammarström. GRF values are based on our own implementation of the method of Pompei et al. (2011); to obtain GQD we use the publicly available qdist software by Thomas Mailund and colleagues.[1] Both programs were tested on the dataset used by Pompei et al. and successfully replicated the results for GRF and GQD that were reported by these authors and cited in Table 7.

[1] Qdist may be downloaded from http://birc.au.dk/software/qdist/.

Table 9. Reticulation and fit between ASJP and expert classifications for Hammarström families (at least 10 members).

Family	δ	*Q*-r	GRF	GQD	*r*	γ	att
Afro-Asiatic	0.3635	0.001643	0.2571	0.0930	0.5803	0.6562	36.52
Algic	0.3826	0.002750	0.3462	0.4343	0.3977	0.5264	36.37
Arawak	0.3527	0.002316	0.3600	0.1782		0.4647	35.64
Atlantic-Congo	0.4210	0.002034	0.4345	0.2555	0.4333	0.3932	34.17
Austroasiatic	0.3556	0.003215	0.2458	0.0564	0.6158	0.6382	37.67
Austronesian	0.3965	0.003406	0.3657	0.2341	0.1205	0.2152	35.12
Carib	0.4274	0.003938	0.6000	0.4418		0.2500	35.28
Central Sudanic	0.2242	0.002970	0.2143	0.0538	0.7744	0.9034	34.10
Chibchan	0.4089	0.001915	0.5294	0.3141	0.5406	0.7000	36.70
Dravidian	0.3225	0.005260	0.4643	0.1285	0.4943	0.6075	34.23
Eyak-Athapaskan-Tlingit	0.3397	0.006016	0.6500	0.2434	0.6015	0.7026	35.30
Heiban	0.3647	0.009270	0.3750	0.1643		0.7963	39.18
Hmong-Mien	0.3161	0.006042	0.2759	0.2204		0.4314	36.88
Ijoid	0.2817	0.007132	0.0000	0.0000		1.0000	36.35
Indo-European	0.2708	0.002194	0.4017	0.0816	0.7946	0.8611	35.75
Je-Kaingang	0.3208	0.008893	0.0000	0.0000		0.9202	34.77
Lakes Plain	0.3585	0.003684	0.1739	0.2096	0.5578	0.7571	35.62
Lower Sepik-Ramu	0.3719	0.002244	0.1333	0.0240	0.5899	0.6842	31.00
Mande	0.2729	0.004468	0.3538	0.1497	0.4664	0.6902	35.19
Mayan	0.2420	0.005362	0.2039	0.0644		0.8273	38.62
Mixe-Zoque	0.2524	0.007532	0.2000	0.0792		0.8831	38.46
Nakh-Dagestanian	0.2483	0.002206	0.0690	0.0179	0.6582	0.9516	39.03
Nilotic	0.2047	0.002665	0.3182	0.0455		0.9283	37.08
North Omotic	0.2688	0.004596	0.0909	0.0741		0.8185	38.56
Otomanguean	0.2651	0.001968	0.0390	0.0009	0.8271	0.9878	37.29
Pama-Nyungan	0.3921	0.002725	0.2843	0.1291		0.6648	33.94
Panoan	0.4000	0.008551	0.8125	0.4993		0.1803	36.58
Quechuan	0.2677	0.011004	0.1316	0.0458		0.5892	38.05
Salishan	0.3019	0.003100	0.0800	0.0677	0.6749	0.8679	38.71
Sepik	0.3374	0.002843	0.0000	0.0000	0.7374	0.9071	34.20

Family	δ	*Q*-r	GRF	GQD	*r*	γ	att
Sino-Tibetan	0.4102	0.002376	0.3582	0.0815	0.5051	0.6551	37.03
Siouan	0.2231	0.002896	0.5000	0.2413		0.9164	36.35
Tai-Kadai	0.2887	0.005999	0.3895	0.1791	0.1601	0.7365	35.04
Torricelli	0.3612	0.001884	0.2609	0.1867	0.4750	0.7240	35.35
Trans New Guinea	0.3942	0.001649	0.1953	0.0558	0.5039	0.6808	35.12
Tucanoan	0.3029	0.008903	0.1786	0.2371		0.7369	36.75
Tungusic	0.3685	0.014542	0.2632	0.2672		0.4012	35.35
Tupi	0.3442	0.005248	0.3256	0.1383	0.7484	0.9113	35.17
Turkic	0.3922	0.012833	0.2653	0.1504		0.4183	36.19
Uralic	0.2788	0.002564	0.1600	0.0498	0.5546	0.9308	38.46
Uto-Aztecan	0.2404	0.003278	0.1290	0.0124	0.9123	0.6983	34.24
West Timor-Alor-Pantar	0.3853	0.004179	0.1351	0.3169	0.2784	0.5223	33.43

Table 10. Spearman correlations for reticulation (δ and *Q*-r) vs. fit with expert classifications (GRF, GQD, *r*, and γ) and correlations with attestations (att) across Hammarström families.

Entities correlated	ϱ	*p*
GRF ~ δ	.3957	.0095
GQD ~ δ	.5551	.0001
r ~ δ	-.6138	.0014
γ ~ δ	-.6992	< .000001
GRF ~ *Q*-r	-.0066	.9671
GQD ~ *Q*-r	.1549	.3274
r ~ *Q*-r	-.1254	.5488
γ ~ *Q*-r	-.1271	.4210
δ ~ att	-3166	.0411
Q-r ~ att	.1075	.4982
GRF ~ att	-.1296	.4133
GQD ~ att	-.1266	.4243
r ~ att	.2131	.3064
γ ~ att	.2818	.0706

Table 10, which provides the relevant correlations, again shows that δ, but not *Q*-r, correlates highly with all four measures of fit between expert classifications and ASJP classifications. This time we observe somewhat higher correlations between att and the other variables, so we also provide each of these. The correlation between δ and att is marginally significant, but none of the measures of fit with expert classifications is significantly correlated with att, so att is at most a mildly confounding factor and does not disturb the overall picture of a solid inverse correlation between the performance of ASJP and the amount of reticulation.

Thus, when discrepancies between our results and those of experts occur, there is a systematic reason that explains much of the discrepancy, namely conflicting phylogenetic signals in the data. This can now be added to the factor of language family size, which was found to correlate negatively with the fit between ASJP and expert classifications in Wichmann et al. (2010a), with Pearson's r being -.44 for *WALS* and -.37 for *Ethnologue*. These correlations are less strong than the ones for δ reported in the present paper.

One cannot conclude from the correlations between differences in classifications and δ that experts are better at dealing with conflicting phylogenetic signals than ASJP, only that they tend to come up with results differing from those of ASJP when there are such conflicting signals. In some cases experts may have been circumventing conflicting signals by arbitrarily assigning languages to certain groups in the absence of good evidence; in other cases they may have been able to identify the causes for reticulation and, after taking into account laterally transmitted traits such as loanwords and diffused phonological changes, may have been able to successfully classify troublesome languages. The latter is certainly not always the case, however. For many of the language families listed in Tables 7 and 9, historical linguistic investigations are scant and superficial. The secure identification of diffused traits within the framework of the traditional comparative method requires the linguist to first have worked out sound correspondences and the phonological developments within lexical items from proto-forms to all modern reflexes, since only then is it possible to identify deviations revealing that a given item has been borrowed. The historical linguist does not start out identifying diffused traits; rather, this is something that belongs to the last and most advanced stage of the reconstruction of linguistic history, and for many of the world's language families this advanced level of research has not yet been reached.

It is beyond the scope of this paper to look at individual cases of discrepancies between ASJP and expert classifications, so we must postpone judgment about the performances of the different methods of classification. We are content to have identified a major cause of discrepancies. The sensitivity of ASJP to real conflicting phylogenetic signals originating in real socio-historical events qualifies it as a tool for directly investigating language contact and linguistic diffusion. In the early days of ASJP (Brown et al., 2008), automated language classification was envisaged as *the* major goal. Having shown how extraordinarily informative the measurement of phylogenetic reticulation is, we can now add the investigation of language contact as an additional goal of ASJP.

5. Conclusions

In this paper we discussed two different measures of reticulation in distance-based phylogenetic data, the δ of Holland et al. (2002) and the more recent *Q*-residuals of Gray et al. (2010). The latter shows sensitivity to the lengths of terminal branches of trees, which should not influence a measure of reticulation, since such branch lengths are due to the distinctiveness and age of taxa, not their fit or lack of fit with a phylogeny. This sensitivity causes *Q*-residuals to strongly correlate with the heterogeneity and age of language families. In contrast, δ shows no such sensitivity. This evidence shows that δ measures reticulation, as it is supposed to, while *Q*-residuals do not. For these reasons we have preferred to use δ for the empirical analyses presented in the paper.

The empirical analyses were directed at trying to discern phenomena that might cause reticulation to occur and to investigate the relationship between reticulation and the performance of ASJP classifications in comparison to classifications by experts. We found that language group size, heterogeneity, and age did not influence δ. Further findings were that languages which, according to standard classifications, are direct offspring of the ultimate ancestor of the phylogeny tend to be more reticulate, whereas there was no trend in the opposite direction for dialects or emerging languages (as defined by designation with the same ISO 639-3 code or by membership in a so-called 'macrolanguage' in *Ethnologue*, respectively). Having established that δ is a realistic measure of reticulation, we went on to compare average δ scores for different language families, showing that these scores explain much of the variance found in the goodness of fit between ASJP and expert classifications.

ASJP has emerged as an efficient tool for pinpointing conflicting phylogenetic signals, and it now remains for future studies to look in more detail at the behaviors of individual languages and pieces of data whose accumulated effects have been traced in these statistical summaries. In particular, we will in the future be interested in studying the effects of loanwords.

Acknowledgements

We are grateful to two anonymous referees for thorough and insightful reports on the first version submitted, as well as to Cecil H. Brown for useful comments.

References

Atkinson, Quentin D., Andrew Meade, Chris Venditti, Simon Greenhill, and Mark Pagel. 2008. Languages evolve in punctuational bursts. *Science* 319: 588.

Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall, and Ingo Plag. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26: 5–42.

Best, D.J. and D.E. Roberts. 1975. Algorithm AS 89: the upper tail probabilities of Spearman's rho. *Applied Statistics* 24: 377–379.

Bowern, Claire. 2010. Historical linguistics in Australia: Trees, networks and their implications. *Philosophical Transactions of the Royal Society B* 365: 3845–3854.

Brown, Cecil H., Eric W. Holman, Søren Wichmann, and Viveka Velupillai. 2008. Automated classification of the world's languages: A description of the method and preliminary results. *STUF—Language Typology and Universals* 61: 285–308.

Bryant, David, Flavia Fillmon, and Russell D. Gray. 2005. Untangling our past: Languages, trees, splits and networks. In Ruth Mace, Clare Holden, and Stephen Shennan (eds.), *The Evolution of Cultural Diversity: A Phylogenetic Approach*, 67–83. London: UCL Press.

Christiansen, Chris, Thomas Mailund, Christian N.S. Pedersen, Martin Randers, and Martin S. Stissing. 2006. Fast calculation of the quartet distance between trees of arbitrary degree. *Algorithms for Molecular Biology* 1:16 (doi: 10.1186/1748-7188-1–16).

Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley, and Stephen C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309: 2072–2075.

Felsenstein, Joseph. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783–791.

Forster, Peter and Alfred Toth. 2003. Toward a phylogenetic chronology of ancient Gaulish, Celtic, and Indo-European. *Proceedings of the National Academy of Sciences of the U.S.A.* 100: 9079–9084 (doi: 10.1073/pnas.1331158100).

Forster, Peter, Alfred Toth, and Hans-Jürgen Bandelt. 1998. Evolutionary network analysis of word lists: Visualising the relationships between Alpine Romance languages. *Journal of Quantitative Linguistics* 5: 174–187.

Gray, Russell D., David Bryant, and Simon Greenhill. 2010. On the shape and fabric of human history. *Philosophical Transactions of the Royal Society B* 365: 3923–3933.

Greenhill, Simon J. 2011. Levenshtein distances fail to identify language relationships accurately. *Computational Linguistics* 37: 689–698 (doi: 10.1162/COLI_a_00073).

Hammarström, Harald. 2010. A full-scale test of the language farming dispersal hypothesis. *Diachronica* 27: 197–213.

Haspelmath, Martin, Matthew S. Dryer, David Gil, and Bernard Comrie (eds.). 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.

Haspelmath, Martin and Uri Tadmor (eds.). 2009. *Loanwords in the World's Languages: A Comparative Handbook*. Berlin/New York: Mouton de Gruyter.

Heggarty, Paul, Warren Maguire, and April McMahon. 2010. Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories. *Philosophical Transactions of the Royal Society B* 365: 3829–3843.

Holden, Clare J. and Russell D. Gray. 2006. Rapid radiation, borrowing and dialect continua in the Bantu languages. In Peter Forster and Colin Renfrew (eds.), *Phylogenetic Methods and the Prehistory of Languages*, 19–31. Cambridge: McDonald Institute for Archaeological Research.

Holland, Barbara R., Katharina T. Huber, Andreas Dress, and Vincent Moulton. 2002. δ plots: A tool for analyzing phylogenetic distance data. *Molecular Biology and Evolution* 19: 2051–2059.

Holman, Eric W., Cecil H. Brown, Søren Wichmann, André Müller, Viveka Velupillai, Harald Hammarström, Hagen Jung, Dik Bakker, Pamela Brown, Oleg Belyaev, Matthias Urban, Robert Mailhammer, Johann-Mattis List, and Dmitry Egorov. 2011. Automated dating of the world's language families based on lexical similarity. *Current Anthropology* 52: 841–875.

Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, and Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica* 42: 331–354.

Huff, Paul and Deryle Lonsdale. 2011. Positing language relationships using ALINE. *Language Dynamics and Change* 1: 128–162.

Huson, Daniel H. and David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.

Huson, Daniel H., Regula Rupp, and Celine Scornavacca. 2010. *Phylogenetic Networks: Concepts, Algorithms and Applications*. Cambridge: Cambridge University Press.

Lewis, M. Paul. 2009. *Ethnologue: Languages of the World*. 16th ed. Dallas: S.I.L. International. Online version: http://www.ethnologue.com/.

McMahon, April, Paul Heggarty, Robert McMahon, and Warren Maguire. 2007. The sound patterns of Englishes: Representing phonetic similarity. *English Language and Linguistics* 11: 113–142.

McMahon, April, Paul Heggarty, Robert McMahon, and Natalia Slaska. 2005. Swadesh sublists and the benefits of borrowing: An Andean case study. *Transactions of the Philological Society* 103: 147–170.

Pagel, Mark. 1999. Inferring the historical patterns of biological evolution. *Nature* 401: 877–884.

Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–290.

Pompei, Simone, Vittorio Loreto, and Francesca Tria. 2011. On the accuracy of language trees. *PLoS ONE* 6: e20109. Downloadable at http://arxiv.org/pdf/1103.4012v1.pdf (accessed April 20, 2012).

R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Downloadable at http://www.R-project.org/.

Tria, Francesca, Emanuele Caglioti, Vittorio Loreto, and Andrea Pagnani. 2010a. A stochastic local search algorithm for distance-based phylogeny reconstruction. *Molecular Biology and Evolution* 27: 2587–2595.

Tria, Francesca, Emanuele Caglioti, Vittorio Loreto, and Andrea Pagnani. 2010b. A stochastic local search approach to language tree reconstruction. *Diachronica* 27: 341–358.

Walker, Robert S., Søren Wichmann, Thomas Mailund, and Curtis J. Atkisson. 2012. Cultural phylogenetics of the Tupi language family in Lowland South America. *PLoS ONE* 7: e35025 (doi: 10.1371/journal.pone.0035025). Downloadable at http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0035025 (accessed April 20, 2012).

Wichmann, Søren. 2010. Internal language classification. In: Silvia Luraghi and Vit Bubenik (eds.), *The Continuum Companion to Historical Linguistics*, 70–86. London/New York: Continuum Books.

Wichmann, Søren, Eric W. Holman, Dik Bakker, and Cecil H. Brown. 2010a. Evaluating linguistic distance measures. *Physica A* 389: 3632–3638 (doi: 10.1016/j.physa.2010.05.011).

Wichmann, Søren, André Müller, Viveka Velupillai, Cecil H. Brown, Eric W. Holman, Pamela Brown, Matthias Urban, Sebastian Sauppe, Oleg Belyaev, Zarina Molochieva, Annkathrin Wett, Dik Bakker, Johann-Mattis List, Dmitry Egorov, Robert Mailhammer, and Helen Geyer. 2010b. The ASJP Database (Version 12). http://email.eva.mpg.de/~wichmann/EarlierWorldTree.htm.

Wichmann, Søren, André Müller, Viveka Velupillai, Annkathrin Wett, Cecil H. Brown, Zarina Molochieva, Sebastian Sauppe, Eric W. Holman, Pamela Brown, Julia Bishoffberger, Dik

Bakker, Johann-Mattis List, Dmitry Egorov, Oleg Belyaev, Matthias Urban, Robert Mailhammer, Helen Geyer, David Beck, Evgenia Korovina, Pattie Epps, Pilar Valenzuela, Anthony Grant, and Harald Hammarström. 2011. The ASJP Database (Version 14). http://email.eva.mpg.de/~wichmann/languages.htm.

Wichmann, Søren and Arpiar Saunders. 2007. How to use typological databases in historical linguistic research. *Diachronica* 24: 373–404.

Wichmann, Søren and Matthias Urban. In press. Towards an automated classification of Englishes. In Terttu Nevalainen and Elizabeth C. Traugott (eds.), *The Oxford Handbook on the History of English*, 977–990. Oxford: Oxford University Press.