

Genetic and Linguistic Affinities between Human Populations in Eurasia and West Africa

DANIEL NETTLE AND LOUISE HARRISS¹

Abstract This study examines the relationship between genetic distance and linguistic affiliation for five regional sets of populations from Eurasia and West Africa. Human genetic and linguistic diversity have been proposed to be generally correlated, either through a direct link, whereby linguistic and genetic affiliations reflect the same past population processes, or an indirect one, where the evolution of the two types of diversity is independent but conditioned by the same geographical factors. By controlling for proximity, indirect correlations due to common geography are eliminated, and any residual relationships found are likely to reflect common linguistic-genetic processes. Clear relationships between genetic distances and linguistic relatedness are detectable in Europe and East and Central Asia, but not in the Middle East, Southeast Asia, or West Africa. We suggest that linguistic and genetic affiliations will only be correlated under specific conditions, such as where there have been large-scale demic diffusions in the last few thousand years, and relative sedentism in the subsequent period.

In a much-cited paper, Cavalli-Sforza and colleagues showed an apparent congruence between a phenogram of allele-frequency distances between human populations and a tree of the world's language families. They propose that this congruence indicates "considerable parallelism between genetic and linguistic evolution" (Cavalli-Sforza et al. 1988:6002). Subsequent studies using diverse scales and methodologies have found variable degrees of association between linguistic and genetic classifications (Barbujani and Sokal 1990; Excoffier et al. 1991; Barbujani and Pilastro 1993; Sajantila et al. 1995; Poloni et al. 1997; Dupanloup et al. 2000; Rosser et al. 2000), depending on the region, the type of diversity measured, and the linguistic classification. Nonetheless, some authors have strengthened their rhetoric to claim "that parallel linguistic allele-frequency change [are] not the exception, but the rule" (Barbujani 1997:1011) or that there is "an intrinsic relation between genetics and language" (Chen et al. 1995:607).

Empirical objections have been raised to the particular linguistic trees used in many of these studies (Bateman et al. 1990; Nichols 1990; McMahon and

¹Department of Biological Sciences, The Open University, Milton Keynes, MK7 6AA, UK.

Human Biology, June 2003, v. 75, no. 3, pp. 331-344.

Copyright © 2003 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: HUMAN POPULATION GENETICS, ALLELE FREQUENCIES, LANGUAGE, EUROPE, ASIA, AFRICA

McMahon 1995). The linguistic tree used in Cavalli-Sforza et al. (1988), for example, contains a set of large clusters drawn from Ruhlen (1987). While about half of these are well-established language families, the other half are speculative entities based mainly on geographical, anthropological, and plausible-guess criteria. Given that these units are not well established on independent linguistic grounds, they can hardly be used to show parallel evolution of genetic and linguistic diversity. The families widely accepted by historical linguists have a drastically different geographic distribution than those posited by Ruhlen (Nichols 1992; Nettle 1999). Furthermore, the linguistic information in most previous studies is quite coarse, relying on the binary parameter of belonging to the same family or not, rather than on discriminating different degrees of relatedness.

There has also been conceptual debate about what a correlation between genetic and linguistic groupings, if demonstrated, would mean (Bateman et al. 1990; Sims-Williams 1998). Of several possible interpretations, the strongest is that linguistic and genetic affiliations reflect the same events in population history. An example would be the migration of founders who reproduce both biologically and linguistically in the new location, and then retain their biological and linguistic identity into the present. Such a strong relationship is assumed, for example, in studies of the peopling of the Americas that use genetic distances as "secondary support for . . . primary inferences based on linguistic . . . data" (Greenberg et al. 1986:486). Models relating Neolithic demographic dispersals to the distribution of language families also rely on the same assumed linkage of genetic and linguistic transmission (Renfrew 1987; Renfrew 1991; Barbujani and Pilastro 1993).

A weaker interpretation is that linguistic and genetic affiliations evolve separately but are conditioned by the same factors. The most obvious of these is geography. Both linguistic and genetic variants have to be transmitted through extensive face-to-face contact, and in the past such interaction relied on close proximity and geographical access. Geographical distance is the best predictor of human population genetic distances (Cavalli-Sforza et al. 1994), and in language families, too, the most closely related languages tend to be situated close together, with relatedness growing more distant as physical distance increases. Thus, both genetic variation and linguistic innovation are subject to isolation-by-distance effects. A correlation with geography is expected in both systems, even if the diversity in the two systems arose at different times and was spread by different events.

The question thus arises whether associations between linguistic and genetic affiliations reflect just the operation of parallel but separate isolation-by-distance processes, or whether there is compelling evidence that languages and genetic variants have actually evolved together. One way to investigate this question is to control for geographic distance and test for a residual relationship or partial correlation between genetic and linguistic affiliations. This test is conservative, since migrations themselves are conditioned by geography. However, like any conservative test it provides the basis for strong inference if it is passed; any

Genetic and Linguistic Affinities in Eurasia and West Africa / 333

residual relationship that is found between linguistic and genetic affiliation can be taken as strong evidence that languages and genetic variants have been spread together.

Sokal (1988) used a partial correlation methodology to show that, for the accepted language families in Europe, there is a relationship between linguistic and genetic affiliations, even when distance is controlled for. However, few studies from regions other than Europe have combined wide geographical scope, appropriate control for geography, and use of rigorous and detailed linguistic classifications. Also, examining data from several genetic systems would be a desirable approach, since a synthetic genetic distance is more likely to show a pattern than any one gene. This paper thus examines the relationships between linguistic and genetic affiliations for populations in five regions of the Old World. Our approach uses only linguistic units that have been unambiguously derived using the strict and independent criteria of historical linguistics, and employs synthetic data from across several genetic systems. We calculate both the unadjusted correlations of linguistic and genetic distances, and the relationships controlling for geographical distance. Any residual relationship that survives these strictures is likely to be a real reflection of tandem transmission of languages and genes.

Methods

Five regions were chosen from the global sample of human populations for which gene frequency data are given in Cavalli-Sforza et al. (1994): Europe (containing 26 populations), West Asia (18 populations), East and Central Asia (21), Southeast Asia (24), and West Africa (13). The populations whose data were used are listed by regional group in the Appendix, along with the language attributed to them and its linguistic classification. The choice of regions was based on the presence of several different well-characterized language families, and the clear identifiability of the geographic location and language spoken by each population. Uncertainties about linguistic classifications in the Americas, Australia, and the Pacific meant that the sample regions all came from the Old World. South Asia was excluded since some of the genetic data coming from that region was from populations based on caste rather than geography or language. Central and southern Africa were also excluded, since the genetic data presented are for population pools rather than populations defined by a single language.

The genetic data are synthesized from population allele frequencies on a large set of classical (non-DNA) nuclear genetic markers. Within each region, genetic distance (F_{ST}) was taken from Cavalli-Sforza et al. (1994) for each pair of populations. Geographic distances in miles between the central points of the ranges of all the populations were estimated by finding the central point of each population. For those populations defined in ethnolinguistic terms a linguistic atlas (Moseley and Asher 1994) was used, and for those that are nation-states or regions a standard atlas was used. The point was estimated as accurately as possi-

ble, usually to the nearest half degree, but sometimes down to the nearest one-tenth of a degree. The great circle distances between all pairs of populations were then calculated using a computer program (Gray 2002). While there is no guarantee that the genetic data were in fact collected at the central point of the population in question, this assumption introduces the least error.

The relationships between the languages of all pairs of populations were classified according to the following numerical scheme: 1, same language; 2, languages in same branch of a family; 3, languages in different branches of same family; or 4, languages not demonstrably related. Only family relationships accepted by the consensus of historical linguists were admitted. The arbiter for linguistic classification was the online version of the *Ethnologue* (Grimes 2002). The branches were taken as the major subfamilies that diverge either simultaneously or in quick succession close to the root of the family tree. For example, for Indo-European, the branches are Armenian, Greek, Indo-Iranian, Slavic, Italic, Germanic, Celtic; for Niger-Congo, Kordofanian, Atlantic, Mande, Benue-Congo, and so on; for Austronesian, Formosan and Malayo-Polynesian (see Appendix for individual classifications). Where a population had lost its language in recent historical times and the original language was known (for example, in the cases of Scotland and Ireland), the original language was used. In all other cases, the language considered was the main or majority one spoken today.

The genetic distances were logged to reduce skewness and kurtosis. First, the unadjusted correlations between genetic distances and geographic and linguistic distances were calculated. The logged genetic distances were then regressed on the matching geographic distances. The residual from the regression line was also calculated. This represents the degree to which the genetic distance is either greater or less than would be expected given the geographic distance. This value was then related to the degree of linguistic affinity. If there is a genuine relationship between linguistic and genetic diversification, then residual genetic distance should increase as linguistic affinity becomes more distant.

Results

Table 1 shows the correlations of linguistic relatedness and genetic distance within each of the five regions. Significant associations occur only in Europe and in East and Central Asia. Linguistic relatedness is significantly associated with geographical distance in all regions except West Africa.

The regression relationships of genetic distance and geographic distance were highly significant in all cases except for West Africa, where the relationship approached significance (Table 2). The equation and r -value for West Africa are in the same range as the others, and the lack of significance is mainly due to the West African sample having fewer populations in it. The equations in Table 2 are surprisingly similar to each other, demonstrating some uniformity in average spatial mobility among individuals across the different parts of the Old World land mass.

Genetic and Linguistic Affinities in Eurasia and West Africa / 335

Table 1. Correlations between the Degree of Linguistic Relatedness and Logged Genetic Distances (Left) and Geographical Distance (Right), for the Five Regions

Region	Genetic Distance ($\ln F_{ST}$)			Geographical Distance		
	<i>r</i>	<i>df</i>	<i>p</i>	<i>r</i>	<i>df</i>	<i>p</i>
Europe	0.48	324	<0.001	0.36	325	<0.001
West Asia	0.07	152	ns	0.29	152	<0.001
East and Central Asia	0.36	135	<0.001	0.30	135	<0.001
Southeast Asia	0.05	252	ns	0.24	252	<0.001
West Africa	0.01	65	ns	0.14	65	ns

Note: *r*, correlation coefficient; *df*, degrees of freedom; *p*, statistical probability; ns, not significant.

Table 2. Regression Equations of Logged Genetic Distance (F_{ST}) on Geographic Distance, for all Pairs of Populations in Each of the Five Regions

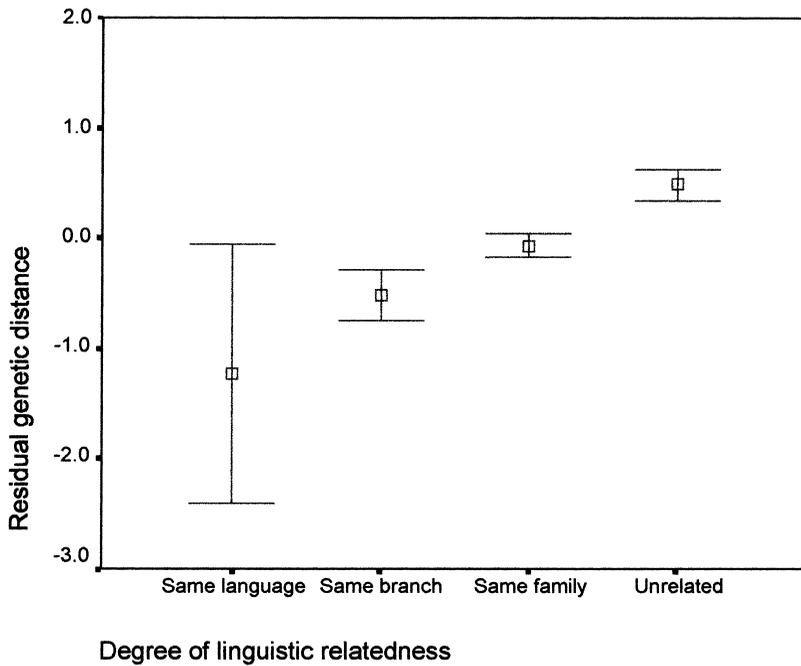
Region	Equation	<i>r</i>	<i>r</i> ²	<i>df</i>	<i>p</i>
Europe	$\ln F_{ST} = 4.07 + 0.00026 \text{ DIST}$	0.36	0.13	323	<0.001
West Asia	$\ln F_{ST} = 4.46 + 0.00054 \text{ DIST}$	0.52	0.27	151	<0.001
East and Central Asia	$\ln F_{ST} = 5.36 + 0.00015 \text{ DIST}$	0.41	0.17	134	<0.001
Southeast Asia	$\ln F_{ST} = 5.63 + 0.00015 \text{ DIST}$	0.22	0.05	251	<0.001
West Africa	$\ln F_{ST} = 4.95 + 0.00017 \text{ DIST}$	0.22	0.05	64	0.079

The residual genetic distances from the equations in Table 2 were sorted by degree of linguistic affinity, and analysis of variance was performed for each region (Table 3). In both the regions where there was a *prima facie* relationship between linguistic affiliation and genetic distance, the relationships survived controlling for geographical distance. The relationships are shown in Figure 1 for Europe and Figure 2 for East and Central Asia. Within these regions, increasingly close linguistic relatedness is associated with decreasing residual genetic distance.

In Europe the pattern is extremely clear; the more closely related two languages are, the closer genetically their populations are relative to what would be expected on the basis of their geographical position. Post-hoc analysis shows that this is entirely due to the genetic structure of the Indo-European speaking populations. Figure 3 breaks down the population comparisons in Europe into those involving two Indo-European languages in the same branch, those involving two Indo-European languages of different branches, those involving two Uralic languages, and those involving two unrelated languages. As the figure shows, Indo-European populations have a systematically close relationship to each other, which is even closer when they belong to the same branch. By comparison, Uralic populations show no genetic affinity at all; they are no closer to each other than

Table 3. Analysis of Variance of Residual Genetic Distance by Degree of Linguistic Relatedness for Each of the Five Regions

<i>Region</i>	<i>F</i>	<i>df</i>	<i>p</i>
Europe	26.87	3,321	<0.001
West Asia	0.74	3,149	not significant
East and Central Asia	3.61	3,132	<0.05
Southeast Asia	1.30	2,250	not significant
West Africa	0.09	3,62	not significant

**Figure 1.** Residual genetic distances by degree of linguistic relatedness—Europe.

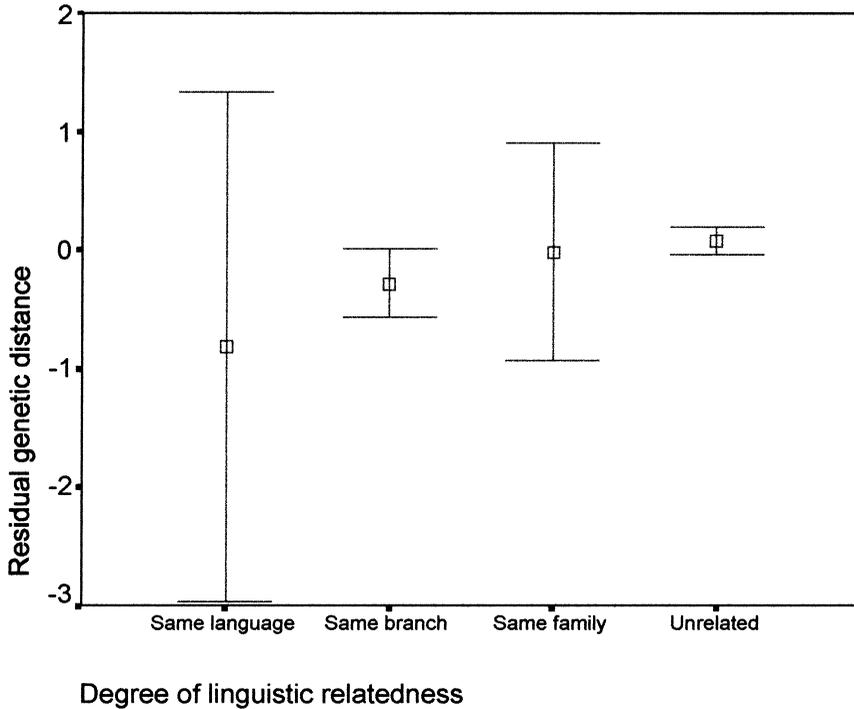


Figure 2. Residual genetic distances by degree of linguistic relatedness—East and Central Asia.

any two unrelated European languages. Thus, Uralic has lost any genetic unity it ever had, and the pattern in Europe is entirely accounted for by the genetic unity of Indo-European, which provides the languages of 22 of the 26 populations in the current sample.

A significant pattern also emerges for the East and Central Asia region, though small numbers of certain types of comparison make the confidence intervals rather large. Post-hoc analysis shows that all three of the main language families in this region show genetic homogeneity relative to unrelated languages (Figure 4). However, there is significant variability, partly due to small numbers of comparisons. Japanese should be considered a single population speaking one language, rather than a language family, but as the figure shows there is significant genetic variability within it.

Discussion

The data show that for Europe, and to a lesser extent for East and Central Asia, there is a detectable association between linguistic affiliation and genetic

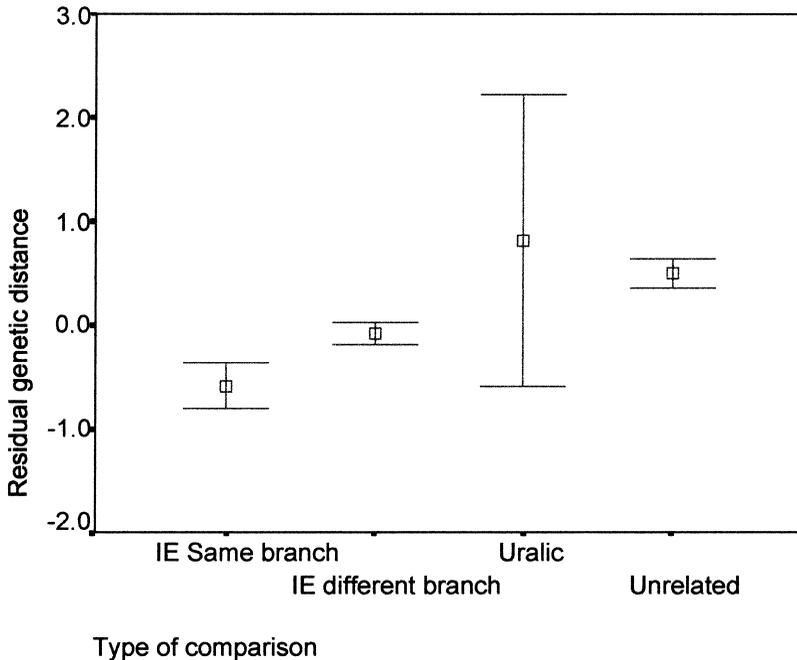


Figure 3. Residual genetic distances in Europe by the language families involved.

distance even once geographical proximity has been taken into consideration. The clarity in the pattern for Europe confirms that the claims of previous studies in this region (Sokal et al. 1989; Barbujani and Sokal 1990; Sajantila et al. 1995) are justified. This is not to say that there are no exceptions, which there clearly are (Sims-Williams 1998), but rather that the overall pattern is discernible through the local complexities. The population history of Europe has long been argued to have involved a major demic diffusion from the East within the last ten thousand years, associated with the spread of agriculture, and *ex hypothesi*, the Indo-European language family with it (Ammerman and Cavalli-Sforza 1984; Renfrew 1987; Renfrew 1991; Barbujani and Bertorelle 2001). The current data support this scenario, and suggest that subsequent local admixture has not been sufficient to completely erase the record of the process. Similarly, in East and Central Asia, the spread of Sino-Tibetan and Altaic protolanguages seem likely to have had substantial demic components whose patterns have not yet been obliterated (Renfrew 1991).

Why similar patterns are not observed in the other regions is not clear. Linguistic relatedness declines with increasing distance in Southeast Asia and West Asia, just as it does in Europe and East and Central Asia, but genetic distances do not covary in the former regions as they do in the latter. No relationship between

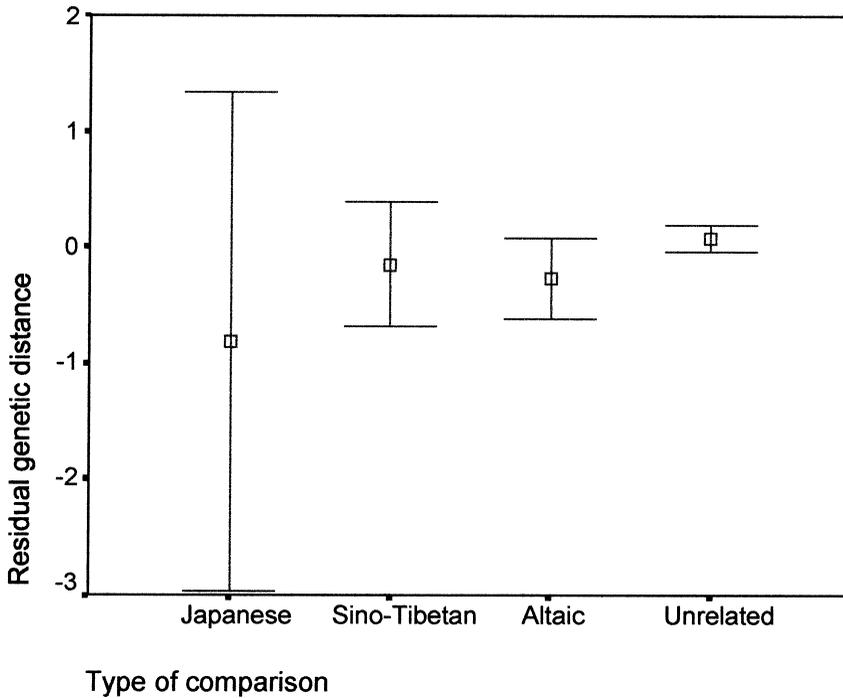
Genetic and Linguistic Affinities in Eurasia and West Africa / 339

Figure 4. Residual genetic distances in East and Central Asia by the language families involved.

linguistic relatedness and genetic distance is apparent, whether geographical distance is controlled for or not.

Southeast Asia has had agriculture-driven linguistic spreads in the last few thousand years—Tai-Kadai and Austronesian. These spreads would have been likely to have a demic component, yet there is no trace of them in the patterns of allele frequency. One confounding variable is that the data set used contains approximately equal numbers of peninsular and island populations. The island populations would be subject to greater local isolation and drift than those on the mainland, which may affect the pattern of genetic distance so much as to lose the signature of the source.

The West African result may simply reflect small sample size, but equally could represent the homogenizing role of pastoral nomadism in this region. Atlantic-speaking Fulani herders have spread throughout the Sahel and intermarried widely with groups speaking languages not just from other branches of Niger-Congo, but from Afroasiatic as well, while nonetheless retaining their language. In the present data set, Nigerian Fulani speakers are genetically much closer to the Hausa farmers among whom they live than they are to their source population, the Senegalese Peul, who speak the same language as they with hardly any dialect

differentiation, but live several hundred miles away. The Fulani spread—the same language found intermingled with language of other families across essentially the whole region—is the probable reason for the noncorrelation of linguistic relatedness and geographical distance in the West African sample. Genetically, too, the gradual diaspora of the Fulani and other herders, which has been a persistent and influential feature of the last thousand years of West African history (Curtin et al. 1995:84–86), would be enough to obliterate traces of prehistoric population affiliations.

Pastoral nomadism would also be a potent factor in explaining the lack of pattern in the West Asia region. Apart from the fastnesses of the Caucasus, much of the region is characterized by nomadism, and its history is one of multiple forays of Indo-European, Afroasiatic, and Altaic speakers over and through each other's paths, no doubt with concomitant gene flow, but in many cases with the retention of distinct languages.

Thus it seems Barbuji's statement that parallel linguistic and allele-frequency change is "not the exception, but the rule" (Barbuji 1997:1011) is too strong. Linguistic relatedness correlates with genetic distances only under certain circumstances—where there are significant and relatively recent demic-linguistic dispersals, and where subsequent local admixture is not so great as to erase their signature. The close correlation of linguistic and genetic diversity may thus be a rarity in global terms. In this study it was found in two out of five regions, but the choice of the Old World as the site of study maximized the likelihood of finding it. In the New World and Oceania, the detection of widespread language family trees has been much more difficult than in the Old World (Nichols 1992; Nettle 1999), probably because those regions lack the large-scale demic-linguistic expansions of the Neolithic. Linguistic relationships thus tend to be more distant and more affected by areal diffusion (Dixon 1997). Under such circumstances, there would seem to be little prospect of discovering the gene-frequency signature of ancient source populations, just as there is little prospect of reducing the linguistic diversity to a few protolanguages, as has been possible for Europe.

The present results show that it is indeed possible to detect the same prehistoric events in the linguistic and gene-frequency diversity of modern humanity, but only in certain places where specific conditions have obtained. The approach can be extended to molecular data sets, where correspondences with linguistic groupings have been less readily observed (Ward et al. 1993; Watson et al. 1996). The positive results that have been put forward using molecular data have tended to use controversial linguistic clusters and coarser grained linguistic classification than the present study (Torroni et al. 1992; Poloni et al. 1997). With more studies that avoid these difficulties, we can move forward from debating the question of whether genetic and linguistic diversity are correlated or not, to a principled understanding of the circumstances under which they match and those under which they do not. In principle, the noncorrelation of linguistic and genetic diversity can be as informative about past population processes as their correlation.

Genetic and Linguistic Affinities in Eurasia and West Africa / 341

Appendix. The Sample Populations by Region, along with the Language Attributed to the Population, and Its Family and Branch

<i>Population</i>	<i>Language</i>	<i>Language Family, Branch</i>
Europe		
Basque	Basque	Basque
Lapp	Sami	Uralic, Finno-Permic
Sardinian	Italian	Indo-European, Italic
Austrian	German	Indo-European, Germanic
Czech	Czech	Indo-European, Slavic
French	French	Indo-European, Italic
German	German	Indo-European, Germanic
Polish	Polish	Indo-European, Slavic
Russian	Russian	Indo-European, Slavic
Swiss	German	Indo-European, Germanic
Belgian	French	Indo-European, Italic
Danish	Danish	Indo-European, Germanic
Dutch	Dutch	Indo-European, Germanic
English	English	Indo-European, Germanic
Icelandic	Icelandic	Indo-European, Germanic
Irish	Irish Gaelic	Indo-European, Celtic
Norwegian	Norwegian	Indo-European, Germanic
Scottish	Scots Gaelic	Indo-European, Celtic
Swedish	Swedish	Indo-European, Germanic
Greek	Greek	Indo-European, Greek
Italian	Italian	Indo-European, Italic
Portuguese	Portuguese	Indo-European, Italic
Spanish	Spanish	Indo-European, Italic
Yugoslavian	Serbo-Croat	Indo-European, Slavic
Finnish	Finnish	Uralic, Finno-Permic
Hungarian	Hungarian	Uralic, Ugric
West Asia		
Pathan Pushtu	Pashto	Indo-European, Iranian
Iranian	Farsi	Indo-European, Iranian
Caspian	e.g. Gilaki	Indo-European, Iranian
Tadzhik	Tadzhik	Indo-European, Iranian
Kurd	Kurdish	Indo-European, Iranian
Kuwaiti	Arabic	Afro-Asiatic, Semitic
Saudi	Arabic	Afro-Asiatic, Semitic
Yemeni	Arabic	Afro-Asiatic, Semitic
Armenian	Armenian	Indo-European, Armenian
North Caucasian	e.g. Chechen	North Caucasian
Svani Georgian	Georgian	South Caucasian
Assyrian	Neo-Aramaic	Afro-Asiatic, Semitic
Bedouin	Arabic	Afro-Asiatic, Semitic
Druse	Arabic	Afro-Asiatic, Semitic
Iraqi	Arabic	Afro-Asiatic, Semitic
Jordanian	Arabic	Afro-Asiatic, Semitic
Lebanese	Arabic	Afro-Asiatic, Semitic
Turk	Turkish	Altaic, Turkic
East and Central Asia		
Ainu	Ainu	Ainu
Hokkaido	Japanese	Japanese
Kyushu	Japanese	Japanese

Appendix. Continued

<i>Population</i>	<i>Language</i>	<i>Language Family, Branch</i>
Ryukyu	Japanese	Japanese
Korean	Korean	Korean
North China	Mandarin	Sino-Tibetan, Chinese
South China	Yue	Sino-Tibetan, Chinese
Nepal no Sherpa	Nepali	Indo-European, Indo-Arya
Sherpa	Sherpa	Sino-Tibetan, Tibeto-Burman
Tibetan	Tibetan	Sino-Tibetan, Tibeto-Burman
Bhutanese	Dzongkha	Sino-Tibetan, Tibeto-Burman
East Uzbek	Uzbek	Altaic, Turkic
Altai	Altai	Altaic, Turkic
Tuva	Tuva	Altaic, Turkic
Yakut	Yakut	Altaic, Turkic
Turk	Turkish	Altaic, Turkic
Turkoman	Turkmen	Altaic, Turkic
Southeast Asia		
Khasi	Khasi	Austro-Asiatic, Mon-Khmer
Mon Khmer	Khmer	Austro-Asiatic, Mon-Khmer
Semai	Semai	Austro-Asiatic, Mon-Khmer
Viet Muong	Viet Muong	Austro-Asiatic, Mon-Khmer
Munda	Munda	Austro-Asiatic, Munda
Chuang	Zhuang	Tai-Kadai, Kam-Tai
Thai	Thai	Tai-Kadai, Kam-Tai
South China	Yue	Sino-Tibetan, Chinese
Bali	Bali	Austronesian, Malayo-Polynesian
Borneo	Dohoi	Austronesian, Malayo-Polynesian
Java	Javanese	Austronesian, Malayo-Polynesian
Sumatra Sunda	Gayo	Austronesian, Malayo-Polynesian
Toba Batak	Batak Toba	Austronesian, Malayo-Polynesian
Malay	Malay	Austronesian, Malayo-Polynesian
Sarawak	Sarawak	Austronesian, Malayo-Polynesian
Sea Dayak	Iban	Austronesian, Malayo-Polynesian
Palau	Paluan	Austronesian, Malayo-Polynesian
Negrito	Agta	Austronesian, Malayo-Polynesian
Philippine	Tagalog	Austronesian, Malayo-Polynesian
Yap	Yapese	Austronesian, Malayo-Polynesian
Atayal	Atayal	Austronesian, Formosan
Bunun	Bunun	Austronesian, Formosan
Paiwan	Paiwan	Austronesian, Formosan
West Africa		
Wolof	Wolof	Niger-Congo, Atlantic
Peul	Fulfulde	Niger-Congo, Atlantic
Serer	Serer	Niger-Congo, Atlantic
Kru	Kru	Niger-Congo, Kru
Mande	Mande	Niger-Congo, Mande
Gur	Gur	Niger-Congo, Gur
Ewe	Ewe	Niger-Congo, Kwa
Yoruba	Yoruba	Niger-Congo, Benue-Congo
Hausa	Hausa	Afro-Asiatic, Chadic
Igbo	Igbo	Niger-Congo, Benue-Congo
Fulani	Fulfulde	Niger-Congo, Atlantic

Received 3 September 2002; revision received 18 February 2003.

Literature Cited

- Ammerman, A.J., and L.L. Cavalli-Sforza. 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton, NJ: Princeton University Press.
- Barbujani, G. 1997. DNA variation and language affinities. *Am. J. Hum. Genet.* 61:1011–1014.
- Barbujani, G., and G. Bertorelle. 2001. Genetics and the population history of Europe. *Proc. Natl. Acad. Sci.* 98:22–25.
- Barbujani, G., and A. Pilastrro. 1993. Genetic evidence on origin and dispersal of human populations speaking languages of the Nostratic macrofamily. *Proc. Natl. Acad. Sci.* 90:4670–4673.
- Barbujani, G., and R.R. Sokal. 1990. Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc. Natl. Acad. Sci.* 87:1816–19.
- Bateman, R., I. Goddard, R. O'Grady et al. 1990. Speaking of forked tongues: The feasibility of reconciling human phylogeny and the history of language. *Current Anthropology* 31:1–24, 175–183.
- Cavalli-Sforza, L.L., P. Menozzi, and A. Piazza. 1994. *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.
- Cavalli-Sforza, L.L., A. Piazza, P. Menozzi et al. 1988. Reconstruction of human evolution: Bringing together genetic, archaeological and linguistic data. *Proc. Natl. Acad. Sci.* 85:6002–6006.
- Chen, J., R.R. Sokal, and M. Ruhlen. 1995. Worldwide analysis of genetic and linguistic relationships of human populations. *Hum. Biol.* 67:595–612.
- Curtin, P., S. Feierman, L. Thompson et al. 1995. *African History: From Earliest Times to Independence*. London, UK: Longman.
- Dixon, R.M.W. 1997. *The Rise and Fall of Languages*. Cambridge, UK: Cambridge University Press.
- Dupanloup, I., A. Schneider, A. Langaney et al. 2000. Inferring the impact of linguistic boundaries on population differentiation: Application to the Afroasiatic-Indo-European case. *Eur. J. Hum. Genet.* 8:750–756.
- Excoffier, L., R.M. Harding, R.R. Sokal et al. 1991. Spatial distribution of Rhesus and GM haplotype frequencies in Sub-Saharan Africa and its relation to linguistic affinities. *Hum. Biol.* 63:273–307.
- Gray, A. 2002. Great Circle Distance Calculator, <http://argray.fateback.com/dist/dodist.html>
- Greenberg, J.H., C.G. Turner, and S.L. Zegura. 1986. The settlement of the Americas—A comparison of the linguistic, dental, and genetic evidence. *Current Anthropology* 27:477–497.
- Grimes, B.F. 2002. *Ethnologue: Languages of the World*. Summer Institute of Linguists www.ethnologue.com
- McMahon, A., and R. McMahon. 1995. Linguistics, genetics and archaeology: Internal and external evidence in the Amerind controversy. *Transactions of the Philological Society* 93:125–225.
- Moseley, C., and R. Asher. 1994. *Atlas of the World's Languages*. London, UK: Routledge.
- Nettle, D. 1999. *Linguistic Diversity*. Oxford, UK: Oxford University Press.
- Nichols, J. 1990. More on human phylogeny and linguistic history. *Current Anthropology* 31:313–314.
- Nichols, J. 1992. *Linguistic Diversity in Space and Time*. Chicago, IL: University of Chicago Press.
- Poloni, E.S., O. Semino, G. Passarino et al. 1997. Human genetic affinities for Y-chromosome P49a/f/TaqI haplotypes show strong correspondence with linguistics. *Am. J. Hum. Genet.* 61:1015–1035.
- Renfrew, C. 1987. *Archaeology and Language*. London, UK: Jonathan Cape.
- Renfrew, C. 1991. Before Babel: Speculations on the origin of linguistic diversity. *Cambridge Archaeological Journal* 1:3–23.
- Rosser, Z.H., T. Zerjal, M.E. Hurles et al. 2000. Y-chromosomal diversity within Europe is clinal and influenced primarily by geography rather than language. *Am. J. Hum. Genet.* 67:1526–1543.

344 / NETTLE AND HARRISS

- Ruhlen, M. 1987. *A Guide to the World's Languages*. Stanford, CA: Stanford University Press.
- Sajantila, A., P. Lahermo, T. Anttinen et al. 1995. Genes and languages in Europe: An analysis of mitochondrial lineages. *Genome Research* 5:42–52.
- Sims-Williams, P. 1998. Genetics, linguistics and prehistory: Thinking big and thinking straight. *Antiquity* 72:505–527.
- Sokal, R.R. 1988. Genetic, geographic and linguistic distances in Europe. *Proc. Natl. Acad. Sci.* 85:1722–1726.
- Sokal, R.R., N.L. Oden, J. Walker et al. 1989. Genetic distances among language families in Europe. *Am. J. Phys. Anthropol.* 76:337–361.
- Torroni, A., T.G. Schurr, C.C. Yang et al. 1992. Native American mtDNA analysis indicates that the Amerind and the Na-Dene populations were founded by two independent migrations. *Genetics* 130:153–162.
- Ward, R.H., A. Redd, D. Valencia et al. 1993. Genetic and linguistic differentiation in the Americas. *Proc. Natl. Acad. Sci.* 90:10663–10667.
- Watson, E., K. Bauer, R. Aman et al. 1996. mtDNA sequence diversity in Africa. *Am. J. Hum. Genet.* 59:437–444.