The Feasibility of Reconciling Human Phylogeny and the History of Language


We address the problem of reconciling human phylogeny and linguistic history and conclude that its resolution requires (1) development of a valid objective method of quantifying linguistic relationships, (2) delimitation and subsequent characterisation of human populations and languages by large-scale demographic census, (3) integration of genetic and linguistic data with other types of information, (4) parallel analyses of the relationships between genetic and linguistic entities using specifically phylogenetic algorithms, and (5) clarification of the biological and philosophical relationship between human lineages and potentially dependent cultural phenomena such as speech. Also, increased discourse between linguists and biologists is needed to distinguish homologous from analogous processes in the two disciplines and thereby standardise terms and concepts. Even if these criteria are eventually satisfied, different processes and rates of evolution and radiation in human populations and languages will continue to complicate attempts to recover "racial" phylogenies. Although profound, these difficulties may not be insuperable, particularly if initial studies sample a regional rather than a global catchment.

RICHARD BATEMAN is Lindemann Research Fellow in the Department of Paleobiology at the Smithsonian Institution [NHB E318, Washington, D.C. 20560, U.S.A.] He was born in 1958 and was Assistant Scientific Officer in the Section for Quaternary Studies of Rothamsted Experimental Station, Harpenden, U.K., from 1977 to 1984. He holds B.Sc.'s in life sciences, from Luton College of Higher Education [1982], and earth sciences, from Birbeck College of London University [1984], and a Ph.D. in Dinantian palaeobotany from London University [1988]. His activities include promoting "demographic" systematics and investigating the roles of saltational macromutation and pollination biology in the megaevolution of higher plants. Among his publications are "A Reappraisal of the British and Irish Dactylorchids" [with I. Denuholm] [Watsonia 14:347-76, 15:321-55, 17:319-49] and "A Reappraisal of the Dinantian Floras at Oxtroad Bay, East Lothian, Scotland" [Transactions of the Royal Society of Edinburgh B, in press].

IVES GODDARD is a Curator in the Department of Anthropology at the Smithsonian Institution, where he has worked since 1976. He was born in 1947 and received his A.B. from Harvard College and his Ph.D. in linguistics from Harvard University in 1969. After a postdoctoral year at the Smithsonian, he taught at Harvard from 1970 to 1976. His research interests are the linguistics and ethnohistory of Native North Americans. His publications include "Algonquian, Wiyot, and Yurok: Proving a Distant Linguistic Relationship," in Linguistics and Anthropology: In Honor of C. F. Voegelin, edited by M. D. Kinkade et al. [Lisse: Peter de Ridder, 1975], Delaware Verbal Morphology [New York: Garland, 1979], and [with K. J. Bragdon] Native Writings in Massachusetts (American Philosophical Society Memoir 185, 1988).

RICHARD O'GRADY has been a postdoctoral fellow in the Smithsonian's Department of Invertebrate Zoology since 1987. He was born in 1958 and received a B.S.c. from the University of British Columbia in 1978, an M.Sc. from McGill University in 1981, and a Ph.D. from the University of British Columbia in 1987. His publications include "Ontogenetic Sequences and the Phylogenetics of Parasitic Flatworm Life Cycles" [Cladistics 1:159-70], [with D. R. Brooks and D. R. Glen] "Phylogenetic Analysis of the Digenea with Comments on Their Adaptive Radiation" [Canadian Journal of Zoology 66:1-41] and "Historical Processes, Evolutionary Explanations, and Problems with Teleology" [Canadian Journal of Zoology 64:1010-20].


RICH MOOI joined the Department of Invertebrate Zoology at the Smithsonian as a postdoctoral fellow in February 1988. Born in 1956 and received his B.Sc. [1981] and M.Sc. [1983] in zoology and his Ph.D. [1987] in evolutionary biology from the University of Toronto. His research interests include the palaeontology and phylogenetic relationships of clypeasteroid sea urchins. Among his publications are "Structure and Function of Clypeasteroid Miliary Spines" [Zoomorphology 106:212-23], "Non-respiratory Podia of Clypeasteroids" [Zoomorphology 106:21-30, 75-90], and "Living and FOssiL Genera of the Clypeasteroida" [Smithsonian Contribution to Zoology, in press].

W. JOHN KRESS is Associate Curator of Botany at the National Museum of Natural History. He was born in 1951 and educated at Queens College, City University of New York, and the American Museum of Natural History, his M.Phil. in 1978 and his Ph.D. in 1986, both in evolutionary biology. His thesis work and subsequent studies as a postdoctoral fellow at the National Museum explored the value of the anatomy of the synax.
for elucidating the phylogenetic relationships of birds. He has more general interests in the development of complex vocal organs and its correlation with complex communication.

The present paper was submitted in final form 5 vn 89.

One reason for our tardiness in treating sociocultural evolution as a selection process is that most of us know a great deal about the vagaries of sociocultural transmission and have an overly simple view of biological transmission.

D.L. HULL, Science as a Process

The instant fame of Cavalli-Sforza et al.'s [1988] "Reconstruction of Human Evolution," which received the rare accolade of rapid review in Science (Lewin 1988a), Nature (Diamond 1988), and Natural History (Gould 1989), reflects the fascination of the question that it raises, namely, whether human phylogeny can be reconstructed by integrating genetic and linguistic information within the broad temporal context provided by the few relevant dated assemblages of fossil humans. We have argued elsewhere [O'Grady et al. 1989] that its main conclusion, that there is "considerable parallelism between genetic and linguistic evolution" [p. 6002], is undermined by analysis of an inadequate database by inappropriate methods and by several conceptual flaws in subsequent interpretations. (For their response, see Cavalli-Sforza et al. 1989.) In this paper, we take the opportunity to discuss our reservations in greater detail, in order to explain how anthropological studies of such potentially profound significance may be made scientifically rigorous.

Constraints on Critical Appraisal

The crux of the paper by Cavalli-Sforza et al. is a phenogram [a tree reflecting overall similarities of entities] that orders "aboriginal" human populations by summary of selected "non-DNA" [i.e., enzyme] allele frequencies using Nei's genetic distance [their fig. 1]. Linguistic "phyla" and "superphyla" [extracted from Ruhlen [1987] and Greenberg [1987]] are superimposed on the phenogram, which we have redrawn as fig. 1.

Critical appraisal of the phenogram (and the paper in general) is hampered by the omission of much crucial information. On five occasions, the reader is referred to additional material, but we have been able to consult) precludes assessment of the rigour of their data collection and analysis.

Delimitation and Characterisation of Genetic Entities

The limited data that are presented by Cavalli-Sforza et al. do not inspire confidence. The data-matrix of 42 entities [putative "racial" groups] × 120 attributes [non-DNA polymorphic alleles] contains 23.7% gaps, reflecting the authors' desire to maximise the number of attributes in the analysis ("The number of genes is of paramount importance for the accuracy of conclusions" [Cavalli-Sforza et al. 1989:1128]). This contravenes the general theoretical principle that no single row of entities or column of attributes in a phenetic data-matrix should have more than 5% missing values [e.g., Temple 1982].

The usefulness of the genetic entities [operational taxonomic units sensu Sneath and Sokal [1973:69]] is questionable. Cavalli-Sforza et al. [1988] repeatedly term these entities "populations," eschewing the term "race" on the grounds that it is "a concept which, for humans, is devoid of any useful scientific definition" [1989:1128]. They delimit all 42 populations a priori, most by partitioning global landmasses into geographic regions and practicing "pooling" within these regions to form geographically delimited groups of organisms [topodemes sensu Gilmour and Gregor [1939]]. They report having "avoided as much as possible the use of individual populations, unless they were unique, and pooled them with others of similar age and used averages [of allele frequencies]. A few unique, isolated populations...were...the average of many samples" [p. 6004]. Hence, the geographic entities termed populations in the data-matrix are aggregates of lesser geographic entities that are also termed populations. Even when 5 European data-matrix populations are further amalgamated, the result is still described as "one population" [p. 6003]. Thus, three different levels of geographically delimited entity are encompassed by the unqualified term "population," and no justification is offered for the geographic locations of the putative population boundaries at any of these levels. Geographic pooling suffices for 36 data-matrix populations, but the remaining 6 [although possessing discrete geographic distributions] are delimited primarily by linguistic affinity (they may be termed "glottodemes"). By opting to use genetic data as attributes, Cavalli-Sforza et al. implicitly assume that these topodemes and glottodemes are also composed of interbreeding individuals (gamodemes) and exhibit significant genotypic differences [genodemes sensu Gilmour and Heslop-Harrison

2. The 16 linguistic entities listed by Cavalli-Sforza et al. are a mixture of "families" and "phyla" as defined by Ruhlen [1987]. Since they treat families and phyla as equal in status, for simplicity we refer to all of these linguistic entities as "phyla."
BATEMAN ET AL. Reconciling Human Phylogeny and Linguistic History | 3

FIG. 1. Phenogram of human populations (1–37) based on average-linkage analysis of Nei's genetic distances representing 120 non-DNA polymorphisms. Linguistic phyla (I–XVI) and superphyla of Ruhlen (1987) are superimposed, and six population aggregates (A–F) recognised by Cavalli-Sforza et al. (1988) are labeled. (Redrawn from Cavalli-Sforza et al. [1988: fig. 1], where it is argued that Melanesians [population 35] possess two linguistic phyla [XIV, XVI]).

[1954]). Thus “population” is used indiscriminately for four fundamentally different types of deme (topodemes, glottodemes, gamodemes, and genodemes), demonstrating that “population” and “race” both require thoughtful qualification when used in a biological context.

What is more serious, subsequent genetic characterisation of the populations sensu lato is essentially typological; samples are too small to permit determination of the range of genetic variation exhibited by the genodemes that they supposedly represent. Although Cavalli-Sforza et al. provide no information on the number (or nature) of individuals analysed per population.3

3. We are unable to penetrate the ambiguities of Cavalli-Sforza et al.’s (1989:1128) subsequent statement that “the average size for gene frequencies is well above 100 individuals. Small samples are rare and were avoided.” Sample sizes should always be reported, together with ranges and/or sample standard deviations if sample size varies.

others (e.g., Cann, Stoneking, and Wilson 1987) have elucidated the numerous difficulties associated with the sampling of human genodemes. Samples are generally small, their geographical coverage local and idiosyncratic, and their genetic “purity” (i.e., absence of admixtures of genes recently acquired from other populations) questionable. Furthermore, different individuals of particular populations tend to be analysed for different attributes by different workers. These variations in sample size, attributes measured, and, especially, population concept represented result in data-sets that are conceptually (though not methodologically) incompatible. In other words, any pair of data-sets sharing attributes can be compared statistically, but such comparison is of doubtful value if the entities represented by the data-sets occupy different levels in the demographic hierarchy (see Bateman and Denholm 1989a, b). We assume that such inconsistencies underlie the “population means”
that comprise Cavalli-Sforza et al.’s basic data-matrix. Certainly, they lacked the large volume of fully compatible data per population necessary to identify the zones of positively correlated depressions in allele frequencies that would allow objective delimitation of genodemes, the most appropriate intraspecific entities for phylogenetic analysis. 4 This is most readily achieved by multivariate ordinations, using individual humans as entities and algorithms that do not require a priori assignment of entities to groups [see Gower 1984]. Once genodemes have been delimited, intragenodeme variation in alleles can be accurately assessed.

The consequences of the typological approach to characterisation of populations would be less deleterious if the resulting mean data-sets were less similar. Cavalli-Sforza et al.’s phenogram is, however, rooted at a small Nei’s genetic distance [0.030], and the most similar pair of populations is linked at a minimal Nei’s genetic distance of only 0.001; given this value, the a posteriori pooling of the European populations could be objectively justified only if the distances among them were all less than 0.001 [i.e., if they were virtually genetically identical]. In this phenogram, 20 of the 26 first-order dichotomies [i.e., those directly subtending at least one population] occur at or below 0.006. We suspect that intrapopulation variation would approximate this figure and that individuals of the same population can therefore be less similar to each other than individuals of different populations in these attributes. This conclusion is supported by the PAUP*-generated parsimony tree for mitochondrial DNA data presented by Cann, Stoneking, and Wilson [1987: fig. 3] [see Spuhler [1988:36–38] for a critical evaluation of this study]. Although individuals from different continents exhibit significantly higher frequencies in certain clusters, overall clustering is poor, reflecting considerable intracontinental variation in gene frequencies. Representation of samples of individuals from a population as a single data-set of mean values for attributes fails to take account of such intrapopulation variation and thus seriously compromises the apparent interpretative value of the phenogram that forms the basis of Cavalli-Sforza et al.’s discussions.

Delimitation and Characterisation of Linguistic Entities

Cavalli-Sforza et al. list 16 of the 17 linguistic “phyla” of Ruhlen [1987], together with two “superphyla” discussed by Ruhlen [1987] and Greenberg [1987]. Ruhlen’s book is a synthesis of both well-established and controversial views. Phyla that are subject to serious criticism, such as Altaic s.l., Austroasiatic, Indo-Pacific, Amerind [sensu Greenberg 1987], and Na-dene s.l. [Ruhlen 1987:130, 148, 180, 221–27], are awarded equal status with generally accepted phyla such as Sino-Tibetan and well-established [and probably relatively recent] phyla such as Indo-European and Dravidian. As Greenberg concedes [in Lewin 1988b:1632], there is little support among specialists for his Amerind hypothesis, which unites all but the northwesternmost Western Hemisphere languages. Criticisms extend throughout the linguistic hierarchy within Amerind, encompassing its six constituent branches (e.g., Northern Amerind), most of its constituent stocks (e.g., Alamosan-Keresioan), the constituent units of some of the stocks (e.g., Alamosan), and even some of the subunits [e.g., Mosan]. Ruhlen’s Austroasiatic phylum [credited to a personal communication from Greenberg [Ruhlen 1987:153]] consists of the Austroasiatic and Austronesian families [a combination originally proposed by Wilhelm Schmidt] plus the Daic [Kam-Tai] and Hmong-Mien [Miao-Yao] families, which lack generally accepted affiliations with each other or with any other families. The original concept of Altaic [comprising Turkic, Mongolian, and Tungusic] is now regarded as discredited; in contrast, an affiliation of Japanese, Korean, and possibly Tungusic is increasingly favoured [Doerfer 1985, Unger n.d., Whitman n.d.]. Other problems may prove more common on detailed investigation. For example, those posed by geographically separate branches of some phyla such as Munda [the branch of Austroasiatic spoken in India] and the forms of Uralic spoken by North Europeans and by very localised languages such as those of Papua (Wurm 1982, Diamond 1988) are obscured by the coarse resolution of the phyla.

Much of the controversy surrounding the linguistic entities can be attributed to substantial variation in the criteria by which they are delimited. To permit valid comparison, both racial and linguistic entities should in principle be delimited objectively and independently. However, although several techniques allow objective delimitation of genetic races, there are no generally accepted analogous techniques in historical linguistics. Perceived linguistic similarities [whether they reflect phonology, grammar, semantics, or syntax] have a highly problematic relationship to historical connections to the extent that they may be less likely to reflect shared inheritance than dissimilarities that can be explained as the results of divergent historical processes [e.g., Hoenigswald 1987:260]. Hence, historical linguists methodologically distinguish convergent [ahistorical stochastic], “genetic” [historically inherited], and diffusional [historically borrowed] resemblances by constructing hypothetical models of the histories of sets of related languages [or, more typically, selected parts of such languages] [e.g., Ruvolo 1987]. The history of language is regarded as a sequence of changes [innovations] between two points in time. Innovations can develop spontaneously within a language or be borrowed from another language. They are considered historically valuable only if they become ubiquitous within a population [e.g., Wiener 1987:218–19].
reconstructed ancestral language (protolanguage) and specifies the innovations that have accrued in each putative descendant via a complex of unique and interdependent historical events.

For example, superficial evaluation of the second-person pronoun prefixes [Sioux] ni-, [Cheyenne] ne-, and [Ojibwa] gi- suggests a match between Sioux and Cheyenne to the exclusion of Ojibwa. However, application of historical-linguistic techniques to the same data demonstrates that Cheyenne ne- and Ojibwa gi- are historically identical. Cheyenne ne- and Ojibwa gi-[in any function] are the expected continuations of Proto-Algonquian *ke- in the respective languages, and the superficial resemblance between Sioux ni- and Cheyenne ne- is a historically spurious accidental convergence.

These innovations illustrate four constraints on quantification:

1. Languages lack the widespread but nonuniversal attributes that characterise major groups of organisms [e.g., the backbones of vertebrates]. The linguistic innovations outlined in n. 7 have no broader application.

2. Once adopted, phonological innovations can be masked by further phonological changes or by functional simplification, termed analogical leveling. Both types of change tend to remove the evidence for the initial phonological innovation and thereby obscure homology [Wieni 1987:222].

3. Homoplasies (parallelisms, convergences, and reversals [e.g., Farris 1983]) are commonplace (e.g., Ruvolo 1987:194; Wang 1987:252) and can covary in greater numbers than do homologues.

4. Languages rarely satisfy the criteria used in biosystematics to establish the direction of evolution (polarity) within a series of related attributes. Polarity assessment, which temporally "roots" a tree, requires that the earlier attributes of a historical series survive unaltered in at least one of the entities being compared. In biosystematics, this polarity assessment is called outgroup analysis (e.g., Farris 1983); its implementation depends not on assumptions of increasing or decreasing complexity of attributes (or of other transformational clines) but on a maximum-parsimony explanation of the extant distribution of the attributes of an evolutionary series. Hypotheses of polarity are tested against one another by the parsimony criterion. Because outgroups can themselves evolve [Wiley 1981], it may not be possible to polarise every attribute in the analysis, and it may be necessary to use more than one outgroup. In evaluating hypotheses of linguistic history, the parsimony criterion is complemented by a considerable body of comparative evidence for the relative probabilities of different types of change, including an evolving understanding of the extent to which general linguistic principles can predict the directionality of changes. Because languages typically do not retain unaltered earlier attributes in sufficient quantities to permit outgroup comparison (e.g., Wiener 1987:220–21), a linguistic tree can be rooted only by untestable a priori assumptions of historical relationships.

Historical linguists therefore regard languages as preconceived conceptual entities that cannot be quantified in isolation; attributes can only be identified and scored by comparison with other putatively related languages. An a priori historical-linguistic phylogeny determines the characteristic states, and a unique line of reasoning supports each postulated event [Hoenigswald 1960, 1973]. Phenetic evaluation of linguistic data will succeed only in straightforward cases of close relationship resulting from equal rates of change. Consequently, the relationships of miscellaneous languages cannot be determined simply by tabulating similar traits.

Thus, techniques developed to classify languages objectively, such as glottochronology and lexicostatistics (e.g., Gudschinsky 1964, Hoenigswald 1967), rely on several controversial a priori assumptions and yield results that are at best ambiguous [Bergsland and Vogt 1962; Ruvolo 1987:194; Rankin 1988:647]. Other attempts to estimate the number or percentage of various types of matches between languages that would rule out chance similarities on probabilistic grounds (e.g., Bender 1969) have not achieved general acceptance. Unfortunately, as is noted by Hoenigswald [1987:57–58], rejection of these techniques precludes objective delimitation of linguistic units of the kind advocated earlier for genetic entities and confines the linguist to typological characterisation followed by analysis using the somewhat tautologous historical-linguistic methodology.

Of the 16 phyla used by Cavalli-Sforza et al. (1967, 1968, 1987, 1988) [Niger-Kordofanian, Nilo-Saharan] are postulated by Greenberg [1963] and Ruhlen [1987] using a phenetic method of classifying languages: "mass comparison" of selected words and grammatical elements of languages and the subsequent tabulation of perceived similarities. This approach assumes that sound correspondences will be lost after a given period following initial divergence [Greenberg 1967:1–37] and produces results similar to those of historical linguistics only at shallow time depths [Goddard 1987, Campbell 1988].

In contrast, the Nostratic superphylum is based on proposed sound correspondences between languages (Kaiser and Shevoroshkin 1988). If the sound correspondences proposed for Nostratic are correct, the premise of Greenberg's method is false, and the Amerind superphylum (and many of its component phyla) must be rejected because no sound correspondences have been demonstrated. Alternatively, if Greenberg is correct, the proposed Nostratic sound correspondences must be spurious; they would merely demonstrate the extent to which accidental similarities between languages can be discovered if the concept of similarity employed is sufficiently broad. Thus, the link between Nostratic and

---

6. These prefixes differ in the particulars of the ways in which they are used but share characteristics such as the inalienable possession of nouns by a second-person singular.

7. The Proto-Algonquian second-person prefix *ke- is directly confirmed by Fox ke-. Proto-Algonquian *e became Ojibwa i and *e became g. Proto-Algonquian *ke- became pre-Cheyenne *ke-, word-initial *k- was lost; *v became Cheyenne n [Goddard 1988]. Sioux ni- is from a similar Proto-Siouan form.
Amerind advocated by Shevoroshkin (1988) and Cavalli-Sforza et al. connects two entities delimited by mutually exclusive (and therefore wholly incompatible) criteria. However, objective partitioning of a phenogram into groups of equal status requires transverse dissection at a specific distance value (e.g., Sneath and Sokal 1973:294–96). By this method, the closest possible approximation to the six groups of Cavalli-Sforza et al. occurs at a genetic distance of 0.011 (fig. 1), but this also requires further partitioning of the African group (San and Ethiopian are separated from the remainder) and the New Guinea/Australia group (New Guinean and Australian populations are separated), thus generating eight population aggregates. The two additional groups are rejected by Cavalli-Sforza et al. (1988: fig. 1) and Gould (1989: unnumbered figure) in favour of more interpretationally convenient but methodologically incorrect resolutions into six and seven groups respectively.

We have attempted to quantify the congruence of the six population aggregates of Cavalli-Sforza et al. (A–F) with the 16 linguistic phyla (I–XVI). By definition, the 7 linguistic phyla awarded parity with single populations (II, III, VI, IX, X, XII, XVI) will agree with any tree. Five of the remaining 9 phyla are confined to one of the six population aggregates each (I to Aggregate A, V to B, VIII to C, XI to D, XIV to E) while the other 4 occur in two aggregate each (IV in A and B, VII in B and C, XIII in C and E, XV in E and F). Thus, the 9 linguistic phyla amenable to a test of agreement at the relatively coarse level of the six population aggregates show 56% (5/9) correspondence.

Correspondence between the genetic and linguistic entities can be assessed at a finer level of resolution by determining whether a linguistic phylum is exclusively shared by inclusive clusters of populations within any one of the six aggregates. Of the five phyla that are confined to a single aggregate, only one (Amerind) corresponds with an inclusive group of all the populations speaking languages of that phylum (XI with 23–25). The other four phyla consist of populations that are not inclusively grouped in the genetic-data tree (I groups 1 and 2, V groups 8–11, VIII groups 15, 17–20, XIV groups 28–35). Thus, correspondence at this more rigorous level of comparison is only 11% (1/9).

Neither of the two linguistic superphyla [Nostratic and Eurasian] precisely corresponds with any of the population aggregates or groups of population aggregates, undermining the “remarkable correspondence” perceived by Cavalli-Sforza et al. (1988:6002) between the superphyla and the North Eurasian group of population aggregates (B–D). The North Eurasian group includes populations that speak non-Nostratic languages (XIII in 16, IX in 21, X in 22, XII in 26), and a Nostratic language occurs in a population [IV in 5] that is excluded from the North Eurasian group. Overall, the Nostratic superphylum encompasses 17% of aggregate A, all of B, and 67% of C. Inclusion of the Amerind phylum within Nostratic [as advocated by Shevoroshkin (1988)] adds 75% of aggregate D. Greenberg’s (1987) Eurasian superphylum largely overlaps Nostratic s.l., encompassing 63% of aggregate B and 89% of C. Cavalli-Sforza et al. (p. 6005) argue that an amalgam of the two superphyla “in-

Congruence of the Genetic Tree and the Linguistic Phyla

In Cavalli-Sforza et al.’s innovative comparison of the genetically based phenogram with Ruhlen’s (1987) linguistic phyla, the phyla are treated as a second set of entities and used as an ostensibly independent assessment of the integrity of the phenogram (though the delimitation of six populations by language effectively precludes objective independence). Unlike the genetic populations, the linguistic phyla are not represented by any quantitative data and therefore cannot be ordered algorithmically to form a tree. Also, most linguistic entities are awarded equal status as “phyla” s.l.; the only exceptions are three subphyla of the Austroycic phylum [Austroasiatic, Daic, Austronesian] and the two largely overlapping superphyla [Nostratic and Eurasian]. Thus, the linguistic “classification” is almost non-hierarchical, and the linguistic phyla can only be superimposed on the genetic phenogram of populations (see fig. 1), a procedure that requires the dubious (and certainly unsubstantiated) assumption that boundaries of linguistic phyla coincide precisely with boundaries of populations or aggregates of populations.

Cavalli-Sforza et al. claim precise parity for seven linguistic phyla and genetic populations, while the remaining nine phyla were associated with aggregates of up to seven populations each. In depicting the relationships of the phyla and populations, Cavalli-Sforza et al. exploit the mobile-like properties of a repeatedly branching tree that are succinctly reiterated by Gould (1989:25); nodes of the phenogram are rotated to achieve maximum apparent congruence of populations and linguistic phyla.

Determination of true congruence is hampered by the non-hierarchical treatment of the linguistic phyla. In these circumstances, congruence is most appropriately assessed by observing whether a particular linguistic phylum corresponds with inclusive clusters of populations on the phenogram. Cavalli-Sforza et al. recognise six population aggregates [Africa, Caucasoid, Northeast Asia, America, Southeast Asia, New Guinea/Australia [A–F in our fig. 1], subtended by nodes that span a relatively broad range of genetic distances (0.008–0.014). Gould (1989) recognises seven aggregates, further dividing Southeast Asia into Southeast Asia sensu stricto and Pacific Islands (node = 0.010). However, objective partitioning of a phenogram into groups of equal status requires transverse dissection at a specific distance value [e.g., Sneath and Sokal 1973:294–96]. By this method, the closest possible approximation to the six groups of Cavalli-Sforza et al. occurs at a genetic distance of 0.011 (fig. 1), but this also requires further partitioning of the African group (San and Ethiopian are separated from the remainder) and the New Guinea/Australia group (New Guinean and Australian populations are separated), thus generating eight population aggregates. The two additional groups are rejected by Cavalli-Sforza et al. (1988: fig. 1) and Gould (1989: unnumbered figure) in favour of more interpretationally convenient but methodologically incorrect resolutions into six and seven groups respectively.

We have attempted to quantify the congruence of the six population aggregates of Cavalli-Sforza et al. (A–F) with the 16 linguistic phyla (I–XVI). By definition, the 7 linguistic phyla awarded parity with single populations (II, III, VI, IX, X, XII, XVI) will agree with any tree. Five of the remaining 9 phyla are confined to one of the six population aggregates each (I to Aggregate A, V to B, VIII to C, XI to D, XIV to E) while the other 4 occur in two aggregate each (IV in A and B, VII in B and C, XIII in C and E, XV in E and F). Thus, the 9 linguistic phyla amenable to a test of agreement at the relatively coarse level of the six population aggregates show 56% (5/9) correspondence.

Correspondence between the genetic and linguistic entities can be assessed at a finer level of resolution by determining whether a linguistic phylum is exclusively shared by inclusive clusters of populations within any one of the six aggregates. Of the five phyla that are confined to a single aggregate, only one (Amerind) corresponds with an inclusive group of all the populations speaking languages of that phylum (XI with 23–25). The other four phyla consist of populations that are not inclusively grouped in the genetic-data tree (I groups 1 and 2, V groups 8–11, VIII groups 15, 17–20, XIV groups 28–35). Thus, correspondence at this more rigorous level of comparison is only 11% (1/9).

Neither of the two linguistic superphyla (Nostratic and Eurasian) precisely corresponds with any of the population aggregates or groups of population aggregates, undermining the "remarkable correspondence" perceived by Cavalli-Sforza et al. (1988:6002) between the superphyla and the North Eurasian group of population aggregates (B–D). The North Eurasian group includes populations that speak non-Nostratic languages (XIII in 16, IX in 21, X in 22, XII in 26), and a Nostratic language occurs in a population [IV in 5] that is excluded from the North Eurasian group. Overall, the Nostratic superphylum encompasses 17% of aggregate A, all of B, and 67% of C. Inclusion of the Amerind phylum within Nostratic [as advocated by Shevoroshkin (1988)] adds 75% of aggregate D. Greenberg's (1987) Eurasian superphylum largely overlaps Nostratic s.l., encompassing 63% of aggregate B and 69% of C. Cavalli-Sforza et al. (p. 6005) argue that an amalgam of the two superphyla "in-

Amerind advocated by Shevoroshkin (1988) and Cavalli-Sforza et al. connects two entities delimited by mutually exclusive (and therefore wholly incompatible) criteria.

Given these differences, we endorse Diamond’s (1988:623) argument that quantitative characterisation of languages would be necessary in order to raise linguistics to the same analytical and interpretative level as the best of the current studies of genetic and morphological data. We recognise, however, that the quantification of language is fraught with hazard, as is the identification of historically significant similarities that are "inherited" within populations rather than diffused among them (e.g., Ruvolo 1987:195; Wiener 1987:222).
includes all, and only, the languages spoken in our major Northeurasian cluster, with the exception of Na-Dene [XII]." In fact, all or parts of two non-Nostratic/Eurasian phyla are included (XIII in 16, XII in 26) and part of one Nostratic/Eurasian language is excluded [IV in 5].

Given 56% correspondence between linguistic phyla and population aggregates at the coarse level of resolution, 11% correspondence at the fine level, and the poor integrity of both superphyla, the parallelism between the genetic and linguistic entities does not strike us as especially "remarkable."

Phenetic and Cladistic Approaches to Phylogenetic Inference

Cavalli-Sforza et al. [1988:6002; 1989] and commentators (Diamond 1988, Lewin 1988a, Gould 1989) assume that the tree of human populations [fig. 1] constitutes a phylogeny that represents a temporal sequence of racial divergences. However, the tree is based on Nei's genetic distance, a measure of overall similarity, and is therefore phenetic (i.e., a phenogram). It does not depict the positions of specific character transitions on the topology of the tree, nor does it distinguish homologous characters derived through descent from a common ancestor [synapomorphies] from those representing retained primitive characters [symplesiomorphies]. Nelson and Platnick (1981) illustrate these concepts with a shark, a lungfish, and a human: although the shark and the lungfish exhibit the greater overall similarity, the distribution of synapomorphies [e.g., lungs] indicates that the lungfish and the human possess a common ancestor not shared by the shark and are therefore more closely related by descent. Tree construction using synapomorphies only is termed cladistics.9 The resulting cladograms are more explicit and more specifically phylogenetic than phenograms [e.g., Wiley 1981, Farris 1983]. The topologies [branching patterns] of the two types of tree can be compared, though a phenogram rarely exhibits the same topology as the most parsimonious cladogram generated from the same data.

In the absence of data, we cannot produce a cladogram for comparison with Cavalli-Sforza et al.'s phenogram. To examine the relationships between the genetic

8. Cavalli-Sforza et al.'s [1989:1128] argument that their tree is "genetic" rather than "phenetic" presumably reflects an erroneous assumption that "phenetic" is synonymous with "phenotypic." Rather than describing a class of data [i.e., genotypic versus phenotypic], "phenetic" describes a class of data-analytical techniques that can be compared with "cladistic" techniques. Phenetic trees reflect overall similarity, their topologies are dictated by branch lengths, and their nodes cannot be related to particular attributes. In contrast, cladistic trees group only by derived homologous similarity [synapomorphy], and their nodes reflect explicit character-state transitions.

Comparative Evolution of Human Genodemes and Languages

Most biologists agree that there is a single history of life (holophyly, the most inclusive form of monophyly), characterised by synapomorphies (shared derived characters) such as the genetic code and the universal molecular symmetry of metabolised sugars (e.g., Margulis 1982). The subsequent history of life is envisaged as a repeatedly branching tree, each branching point representing a speciation event. The topologies of most portions of this "tree of life" remain highly speculative, as does the degree of constancy of overall evolutionary rate through time (e.g., Kimura 1983, Levinton 1988). Organic evolution is mediated and constrained by the extremely conservative intrinsic mechanism of genetic inheritance. Gene mutations (duplication, addition, deletion, inversion, transposition, and substitution events), even if they confer great adaptive advantage, can only spread gradually through genodemes over many generations.

Superficially, analogies between the evolution of life and that of language are attractive. The possession of language per se and the evolution of particular languages both appear amenable to explanation in terms of adaptive fitness. Also, languages clearly undergo "descent with modification," manifested as changes in the shapes and arrangements of words and other elements that are analogous to genetic mutations. However, unlike genes, words are not intrinsically constrained to the reproductive cycle via the genome. Like behavioural traits, they can be acquired during the lifetime of the individual; transcriptional modifications are both more readily...
achieved (either consciously or subconsciously) and more readily rectified. As with mutant genes, the success of the linguistic novelty is most appropriately measured by its proliferation.

In contrast with genes, however, linguistic "mutations," because they are acquired, can be disseminated throughout a human population in a single generation. Consequently, as is noted by Diamond [1988:622], there are many recent examples of the transfer of languages with little associated transfer of genes. Two or more languages can be maintained concurrently, both by an individual and by a society, though in most circumstances one language competitively displaces the other(s). Moreover, languages need not be transmitted as complete sets of words, analogous to genes in a gamete; each word can function as an independent entity. This characteristic allows borrowing of linguistic elements, a process similar to biological hybridisation. As in biology, geographical proximity of the parental stocks aids this hybridisation [Hoenigswald 1987:262–63], but unlike biological entities, closely related languages are not inherently more likely to hybridise [cf. Stace 1975; Wiener 1987:225]. One of the few parallels between the evolutionary constraints on human populations and those on languages that withstands close scrutiny is the shared lack of intrinsic barriers to hybridisation and consequent reliance on extrinsic (geographical and, especially, behavioural) isolation. Whereas biological hybridisation inevitably involves the transfer of a complete set of gene-bearing chromosomes, hybridisation of languages tends to be confined to specific elements and therefore has more appropriate biological analogues in "jumping genes" [e.g., Campbell 1983] and virus-mediated gene transfer than in hybridisation of species. Opinions differ on the extent to which some words and concepts constitute an inner core that is much less prone to being borrowed and therefore much more likely to contain retentions [cf. Ruvolo 1987, Wang 1987]. Although borrowing generates homoplasies throughout language as a whole, it does create innovations within particular linguistic lineages. Hybrid words or, more commonly, phonological elements [mergers] can also function as innovations in speech communities [Hoenigswald 1987]. However, large-scale borrowing can obscure relationships of languages, especially if the borrowing is reciprocal. Elements of different parental languages may then become difficult to distinguish, reflecting a process analogous to genetic introgression [e.g., Stace 1975:3]. Borrowing and convergences perturb hypothetical, neatly dichotomous linguistic phylogenies and create anastomoses [Wang 1987], certainly at the level of linguistic elements and possibly at the level of languages themselves [Thomason and Kaufman 1988].

To summarise, most evolutionary processes in languages appear fundamentally different from their analogues in living organisms. Hence, evolutionary patterns may also be different, and the congruence between human genodemes and languages sought by Cavalli-Sforza et al. may be an unreasonable expectation.

Any primary congruence has been further disrupted by language loss and migration, factors that must have had a profound cumulative effect throughout human history. The existence of language isolates (each effectively a monotypic phylum) also complicates resolution of the prehistory of the world's languages. Ruhlen [1987] recognises only 9 isolates that cannot be accommodated in his 17 "families and phyla" [Basque, Burushaski, Etruscan, Gilyak, Hurrian, Ket, Meroitic, Nahali, and Sumerian], though many other linguists would considerably expand this list. These languages are either spoken by single isolated populations or long extinct and are probably the remnants of a much larger number of extinct (and therefore largely undocumented) languages that would have been spoken by members (or forebears) of the 42 putative extant populations but would not have been assignable to any of the 17 extant phyla.

The argument that language replacement in Europe was exceptional, reflecting the emergence of elites during the last 5,000 years only [Renfrew 1987; Cavalli-Sforza et al. 1988:6005; Diamond 1988:622], is weakened by the existence of widespread shallow linguistic groupings in band-level societies such as Inuit-Inupiaq [Beringian Diomede Islands—East Greenland [McGhee 1984, Woodbury 1984]], Pama-Nyungan [most of Australia [Dixon 1980]], and Numic [Great Basin of North America [Miller 1986]]. Moreover, if the recent distribution of languages does partly reflect dissemination by elites, correlation between languages and populations cannot be expected. For example, Cavalli-Sforza et al.'s European population includes speakers of Basque, a non-Indo-European language, and speakers of three branches of Indo-European (Germanic, Italic, and Hellenic). Since the spread of Indo-European languages across Europe presumably postdated the establishment of the European population, available data suggest that Basque rather than Indo-European should be regarded as the original language of the European population. Even if only the Indo-European-speakers in Europe are considered, correlation between the branches of Indo-European and any putative populations within Europe is poor. Furthermore, the English component of the European population has shifted its language twice within recent millennia, thus further obscuring recognition of an "original" European language.

Similarly, the present languages of Australia probably originated much later than the date of 40,000 years B.P. estimated for the earliest human occupation [Dixon 1980:19]; any distinct genetic and linguistic characteristics presumably largely reflect the relative geographical isolation of the island continent. Where geography constrains migration to a narrow route (e.g., through the

---

12. Dixon (1980) has suggested that Pama-Nyungan may not be a bona fide linguistic entity, as some languages of divergent types may be derived from the otherwise rather uniform Pama-Nyungan languages. If he is correct, the point made here is reinforced.

Beringian bottleneck between Asia and North America, character distributions are likely to be clinal. In both cases, language groupings are considered more recent than the human occupation of the areas in which they are spoken. Furthermore, Takahata and Nei [1985] and Spuhler [1988:24] argue that gene divergence appreciably pre-dates population [i.e., racial] divergence, suggesting a consistent progressive order of temporally distinct events: initial gene divergence → population divergence → linguistic divergence. Increasing the time lags between events decreases the likelihood of generating congruent phylogenetic and linguistic trees.

Establishing a historical correlation of phylogeny and linguistic relationships presupposes the ability to establish the absolute time depth of both phylogenetic and linguistic divergences. Unfortunately, even if the current speculative divergence dates for the human populations are accepted, there is no generally accepted comparable method of dating linguistic divergences.

The physical ability to speak is generally regarded as a monophyletic trait (e.g., Greenberg 1987), though monophyly cannot be assumed for language itself (Wiener 1987:220). Mellars [1988] and Cavalli-Sforza et al. [1988, 1989] argue that human language emerged coevally with modern man ([Homo sapiens] subsp. sapiens) ca. 100,000 years B.P., on the assumption that language as we know it emerged at the same time as the only subspecies known to employ it. Laitman (1985), however, concludes that the earliest H. sapiens (ca. 300,000 years B.P.) possessed the anatomy necessary for human speech. If, as he suggests, the evolution of the organs of speech conferred selective disadvantages [notably greatly increased vulnerability to choking], these must have been outweighed by even greater selective advantages. It could be argued that adaptation for another biological function preadapted vocal organs for subsequent linguistic use [thus increasing the likelihood that speech per se is polyphyletic], but a convincing alternative role has not been proposed (see Davidson and Noble [1989] for speculations concerning the origin of language).

We agree with Cavalli-Sforza et al. [1989] that a large temporal gap probably separates the evolution of the ability to speak and the emergence of what we now recognise as distinct languages. If so, divergence of human genodemes pre-dated the appearance of languages in these populations and language is, by definition, polyphyletic (fig. 3). This would not preclude the construction of a tree connecting the independently evolved languages such as that of Cavalli-Sforza et al. but would render attempts to discover historical relationships between these unrelated entities entirely artificial. Even if adaptation in early H. sapiens specifically for speech is
likely, the implied time depth for the origin of language would place most of the history of the world's languages far beyond the reach of available analytical techniques. Many divergences are probably too remote to reconstruct their shared protolanguage, thus restricting interpretation to sister-group relationships (e.g., Platnick and Cameron 1977, Farris 1983), which are very difficult to assess in the absence of a temporal framework.

One of the most important factors underpinning comparison of the evolution of human populations and languages is constancy of rate of change. Although the Nei-type genetic phenogram (fig. 1) is interpreted phylogenetically by Cavalli-Sforza et al. (1988, 1989), Diamond [1988], Lewin [1988a], and Gould [1989], they recognize that it is dependent on a general assumption of constant evolutionary rates in human populations. Unfortunately, Nei's genetic distance is non-metric and therefore cannot generate a phenogram with branch lengths that would accurately reflect constancy of rate [Farris 1981; Ruvolo 1987:200]. Moreover, the assumption of rate constancy remains controversial (cf. Farris 1981; Kimura 1983; Gingerich 1986; Spuhler 1988:23), especially when the genetic entities being compared are not species but local or regional populations (gamodemes) lacking intrinsic barriers to gene flow. (Both Cavalli-Sforza et al. [1988] and Gould [1989] selectively reject the constant-rate hypothesis for certain postulated racial divergences on the phenogram that are contradicted by migration hypotheses based on dated fossil assemblages.) The lack of intrinsic controls on the evolution of languages has undoubtedly precluded the development of a "linguistic clock"; rates of change in languages cannot be assessed retrospectively.

Trees for both human populations and languages will lack most ancestral stocks, whose presence cannot be assumed. Many extinctions of both races and languages undoubtedly occurred. Thus, if repeatedly dichotomous trees are a reasonable representation of the patterns of evolution in human populations and languages, they should be constructed using algorithms that are independent of time and focus on topologies reflecting character transitions at nodes rather than phenetically determined lengths of branches. The application of such cladistic algorithms would require the identification of inherited homologous characters (preferably including autapomorphies for each entity) and their subsequent polarisation into transformation series of increasingly derived character states (e.g., Platnick and Cameron 1977, Farris 1983). Difficulties inherent in applying these techniques to genetic (e.g., Dunn and Everitt 1982:131–37) and linguistic (e.g., Goddard 1975:255) data will only be overcome if they are seriously addressed.

Replicators and Interactors, Entities and Attributes

At a deeper level, the issues raised by Cavalli-Sforza et al. reflect one of the most fundamental controversies in evolutionary biology: the identification of units of selection and their role in evolution (e.g., Hull 1988:chap. 11). Dawkins [1976] has distinguished between the narrow concept of molecular genes and the broader concept of evolutionary genes by terming the latter "replicators." A replicator has structure, longevity, fecundity, and fidelity and is defined by Hull [1988:408] as "an entity that passes on its structure largely intact in successive replications." It is contrasted with the concept of "interactor" (Dawkins' [1982a] "vehicle")", an entity that interacts as a cohesive whole with its environment in such a way that this interaction causes replication to be differential" (p. 408). Hull then defines selection as "a process in which the differential extinction and proliferation of interactors cause the differential perpetuation of the relevant replicators" (p. 409). Replicators and interactors are types of wholly interdependent entity that share many properties. Both are individuals (spatiotemporal particulars) of finite duration and cohesive to the extent that they are incapable of indefinite change through time. Both also have active roles in evolution, though only interactors function in the selection process. Neither is restricted to a specific level of biological organisation.

In complex, allogamous species such as man, the most obvious replicators are genes (e.g., Dawkins 1976), and the most obvious interactors are individual organisms (e.g., Hull 1988), though some biologists view entities lower (e.g., genes [Dawkins 1982a, b]) or higher (e.g., avatars) in the bio-organisational hierarchy as the prime interactors. As replicators produce sequences of themselves through time (faithfully or with modification, with or without interaction), they generate a lineage, a third type of entity that "persists indefinitely through time, either in the same or an altered state, as a result of replication" (Hull 1988:409). A lineage is sufficiently incohesive to change indefinitely but sufficiently cohesive to maintain its historical continuity.

These three concepts of biological entity (replicator, interactor, lineage) provide a context for the similarly fundamental (though rarely critically analysed) conceptual distinction between entity and attribute (cf. Temple 1982). Most scientific research involves investigating the relationships between entities by recording their attributes. In a data-matrix, the entities are represented by data-sets of values for one or more (typically many

14. The term "vehicle" is ambiguous; it has been applied to other concepts in linguistics (Richards 1936:56) and to non-biological replicators in conceptual evolution (Campbell 1979; Hull 1988:414).
15. Damuth's (1985) avatars are groups of organisms of a single species in a single geographically restricted community and therefore encapsulate three deme concepts: ecodeme, topodeme, and genodeme (see Gilmour and Heslop-Harrison 1954).
16. Eldredge (1989:140) preferred to award primacy to reproduction ("moremaking of entities of like kind") rather than replication, arguing that "both interaction and reproduction must occur within a single category of entity at a given level [in the bio-organisational hierarchy] for selection truly analogous to natural selection to be said to occur at levels higher (or lower) than the organismic level."
more] potentially shared attributes. The only constraint on an attribute is that it be less inclusive than the entity that it describes. This characteristic has two important implications. First, it offers the scientist considerable flexibility of approach within the organisational hierarchy of entities; for example, the same attribute can be recorded for an individual organism, a deme, or a species. Second, an attribute can be defined only relative to a particular entity. A systematist may use morphological features as attributes to characterise species-level entities, but an ecologist may use the resulting species as attributes of more inclusive communities such as species. Thus, an attribute is not conceptually distinct from an entity; rather, it is a temporary methodological status awarded to entities that are used to investigate the relationships of other, more inclusive, entities.\footnote{This conclusion renders rather less profound Hull's (1988:523) argument that “conceptual entities are replicators, not traits” (i.e., not attributes).}

Although they are complex and difficult to convey, we believe that these philosophical concepts provide vital insights into whether evolutionary patterns and processes in human populations are homologous or merely analogous and therefore whether or not they are potentially reconcilable.

The phenogram presented by Cavalli-Sforza et al. ostensibly depicts relationships between lineages of human populations, using overall similarities derived from the quantification of selected genes. The probable prime replicators (individual genes) are used to assess the historical relationships of aggregates of the probable prime interactors (individual humans). The status assigned to the linguistic phyla is much more ambiguous. Cavalli-Sforza et al. present human populations and linguistic phyla as separate lists of entities that can be directly compared, our discussion of the correspondence of population clusters and linguistic phyla makes the same tacit assumption. However, in order to calculate a consistency index for their relationship, we treat the languages as dependent attributes of the populations, potentially capable of resolving their phylogenetic relationships.

Languages undoubtedly form lineages, persisting indefinitely through time as a result of replication. A language also possesses the characteristics of a replicator (structure, longevity, fecundity, and fidelity), and its structure persists through successive replications. However, unlike a gene, a language is not pre-programmed for replication, and its role during replication is passive; it is propagated through the conscious will of its human host. Prior to the advent of writing, languages had no independent existence. A language is also dependent upon its host for its expression and so cannot function as an independent interactor with its environment to cause differential replication. Languages exhibit differential replication (i.e., undergo selection) only as a result of interactions between humans, either as individuals or as societies (i.e., speech communities). Such cultural replicators have been termed “memes” by Dawkins (1976:206), who has promoted the controversial argument that they are features of an extended concept of phenotype (Dawkins 1982a, b). If so, languages are merely attributes of humans and can in theory be used to infer human relationships, either independently or by comparison with patterns exhibited by other types of attribute. Although the possession of speech is reflected anatomically and confers strong selective advantage, the possession of a specific language is much more prone to conscious manipulation. Human languages represent conceptual systems that can be selected (in their entirety or in part) and adapted by particular groups of humans. The selective advantage in possessing language per se is generally much greater than for other conceptual systems such as religious or political allegiances, but the adoption of a particular language is determined by cultural pressures similar to those that determine adoption of a particular religion or political stance. Thus, it is more difficult to regard a human language as a phenotypic character than to so regard, for example, pair-bonding in birds, which may be at least partly inherited. We doubt that a Frenchman is born with a genetically induced predilection for the French language (cf. Wang 1987).

Also, as is noted by Cavalli-Sforza and Feldman (1981), Wiener (1987:219), and Wang (1987), language transmission has a vertical [parental, phylogenetic] component, but this may be subordinate to oblique (non-parental teacher) and horizontal (peer-group) components.

Because languages do not reside in the genome, they are freed from the severe constraints that restrict the occurrence of genetic innovations and the speed of their subsequent dissemination through biological populations. Most important, attributes such as language are masked from direct selection and selected for indirectly by factors such as the social status of their proponents (Wien 1987:219, 225). As a particular language is not subject to direct selection, it need not be adaptively superior to other languages to increase in frequency (witness the present relative frequencies of an eclectic, ad hoc historical language, English, and a deliberately manufactured “universal” [and ostensibly ahistorical] language, Esperanto). Indeed, it need not compete directly with other languages (in the way that alternative alleles compete for occupancy of a genetic locus), since a single human can assimilate several languages. Consequently, we believe that the success of a language is primarily determined not by the genealogy of its advocates but by their social influence, assisted by positive feedback that tends to sustain the expansion of a language whose frequency is already increasing.

We conclude that human languages “evolve” only in the most general sense. Many of the superficially attractive similarities between patterns of genetic and linguistic inheritance reflect analogous rather than homologous processes.

Conclusions

Satisfactory reconciliation of human phylogeny and linguistic history requires several methodological and conceptual advances:
1. Increased discourse between linguists and biologists is needed. Distinguishing homologous and analogous processes in linguistic and biological evolution would allow standardisation of terms and concepts where appropriate.

2. Objective techniques for quantifying languages are desirable in order to provide data for algorithmic analysis. Admittedly, some linguists doubt that such techniques can be successfully applied.

3. Human populations and languages should be delimited by large-scale demographic census and subsequent multivariate ordination. This should reveal any positively correlated depressions in frequencies of character states that would reflect meaningful boundaries of entities and thereby minimise the effects of the problems associated with partitioning continuous variation (e.g., Prentice 1986).

4. The resulting entities could then be characterised using large quantities of fully compatible data collected throughout their geographic and ethnographic ranges. Extant hominids are sufficiently abundant and widespread to allow large-scale demographic sampling. Although the paucity of fossil hominids restricts palaeoanthropological studies to small-scale typological characterisation, recent technical advances have allowed DNA studies of fossil vertebrates (e.g., Pääbo 1989, Pääbo, Higuchi, and Wilson 1989), thereby providing a more direct method of tracing the evolution of human populations.

5. The laudable interdisciplinary approach of Cavalli-Sforza et al. to the integration of genetic and linguistic data should be expanded to include other types of relevant information. For example, although previous attempts at morphological analysis of humans have proved controversial and often been discredited (Gould 1981), there is no scientific reason for abandoning the collection of such data.

6. Phylogenetic interpretation is facilitated by the application of inherently historical (i.e., cladistic) algorithms. They generate phylogenies that can be related to specific character transitions and are not dependent on assumptions of constant evolutionary rates. All types of data can be summarised in a single phylogeny, or several phylogenies can be compared a posteriori. However, it may be impossible to determine true congruence between such phylogenies, as the boundaries of different types of entity (e.g., human genodemes versus languages) are unlikely to coincide precisely and the degree of interdependence between organisms and cultural traits such as language and behaviour requires clarification.

7. Such phylogenies are likely to be "noisy," reflecting complicating factors such as hybridisation, the less constrained evolution and competitive displacement among cultural traits, and the fact that contemporaneity of human genodemes and languages cannot be assumed. It could be argued that relatively recent languages are exceptionally complex and therefore provide a poor analogy for the simpler patterns that presumably characterised the early evolution of language [particularly if dominated by iconic [representational] rather than symbolic [arbitrary] grammatical processes [Swadesh 1971: 115–56]]. Any benefits gained from simplification of linguistic patterns with increasing time depth would, however, be negated by constraints imposed by lack of documentation and inferior time resolution.

8. The demographic methods that we advocate for delimiting languages and human races are essentially agglomerative, using individual humans as basic operational taxonomic units. They therefore require acquisition of substantial bodies of compatible data and preclude a rapid solution to the global problems posed by Cavalli-Sforza et al. We suggest that more intense, geographically localised sampling confined to regions of greatest potential offers a much better opportunity to reconcile some population and linguistic divergences. Promising candidates include the Beringian region (Krauss 1988), New Guinea (Wurm 1982, Stoneking, Bhatia, and Wilson 1986), Australia (Dixon 1980), and China (Yu et al. 1988). Experience and conceptual progress achieved during investigations would increase the probability of success in even more challenging regions such as Europe, central Asia, and Melanesia. Nevertheless, satisfactory global resolutions remain a very distant goal.

Comments

DAVID F. ARMSTRONG

Gallaudet University, Washington, D.C. 20002, U.S.A.

Bateman et al. provide a thorough critique of the combined linguistic/phylogenetic approach to analyzing human diversification proposed by Cavalli-Sforza et al. Their critique could, however, be extended a step further to a full exploration of the question when and under what conditions the application of tree diagrams to human subspecific variation is justified. Dichotomous tree diagrams may not be the most appropriate models for describing either the genetic or the linguistic histories of large proportions of the world's current and former populations.

Cladograms are most appropriate to the description of populations that are known to be diverging or that have diverged genetically, that is, at the species level and above (e.g., Rensch 1959). It is arguable whether they may also be appropriately applied at the subspecific level to populations that experience extremely limited gene flow over long periods of time (see n. 4). Humans are the most mobile and adaptable of the mammals, and it is unlikely that any pair of modern human populations is or was headed toward intergroup infertility or speciation.

For example, according to the scheme of Cavalli-Sforza et al., the population of modern Ethiopia represents an anomaly—genetically African but speaking a language that belongs to a linguistic group [Afro-Asiatic] whose speakers are primarily "Caucasoid." The tree diagram showing a phylogenetic split between African and
other peoples forces Cavalli-Sforza et al. to describe Ethiopians as a primarily African population with "genetic admixture." The historical model this implies is that, after the split, invading Eurasians came back to Africa bringing their genes and languages. There is another possible historical model that the tree diagram cannot account for—that there never was a split sufficiently complete to justify representation of a phylogenetic branch and that, since the expansion of modern humans out of Africa, Ethiopians and other inhabitants of Northern Africa have been clinally intermediate between Sub-Saharan Africans and Eurasians. This model would not preclude the back-migration from Eurasia to Africa—it would accommodate it easily. Ethiopians, on the African side of the Red Sea, are, not surprisingly, genetically more like Africans from further south on the continent, and Arabs, on the Eurasian side, are more like other Eurasians. In fact, the question is open whether major populations of Homo sapiens sapiens have ever been separated to the extent and over the periods of time that would be required to justify the representation of their current diversity or history through the use of tree diagrams, whether phenetic or cladistic. This, along with recognition of the need for vigilance against racist interpretations, was the fundamental observation that underlay the anthropological attack on the race concept that was mounted a quarter-century ago (e.g., Montagu 1964).

The question is not whether trees can be fitted statistically to the data but whether they represent the most appropriate models. Bateman et al. point out that anastomoses of linguistic stocks call into question the application of tree diagrams to linguistic diversification. (The same can equally be said of genetically defined populations.) The origin of English provides a familiar example of the problem. It is common to place modern English at the end of a branch of the Germanic group of Indo-European languages, because of the known history of its speakers and elements of its core vocabulary and sound system. It could be argued, however, on the basis of lexical and syntactic features, that it is really Italic and not Germanic, closely related to modern French. The truth, of course, is that it is neither—it is descended from an amalgam of Anglo-Saxon and Norman French and has evolved subsequently for a millennium relatively independently of other Germanic languages. Gould (1989:20), in his review of Cavalli-Sforza et al., recognizes this point but does not discuss its implications for the representation of linguistic evolution by constantly branching tree diagrams. The tendency to place modern English and French at the ends of two ramified branches having a common ancestor as remote as Proto-Indo-European grossly distorts the actual relationship between the two languages and their speakers, and the situation of English is not unique in human history.

Bateman et al. recognize and discuss the implications of these and other problems inherent in the use of tree diagrams, but more discussion of a philosophical nature is needed, extending beyond the adequacy of statistical measures of genetic distance and "treeness." There may be more appropriate theoretical structures that could be constructed to describe the genetic and linguistic diversification of human populations, for example, a reticular structure stretched over a sphere (globe) and growing outward through time. Cavalli-Sforza et al. (1988:6005), referring to Wolpoff, Wu, and Thorne (1984), recognize the possibility that continuous network models could be applied to these data but dismiss it. Cavalli-Sforza has been producing human phylogenetic trees, based on merged sets of diverse data, for many years, and his efforts continue to be valuable. We should welcome attempts to rejoin phylogenetic and linguistic data in nonracist formulations, although this approach has been in disrepute for much of this century. However, despite increases in the quantity and quality of data and the sophistication of methods, questions raised previously (Barnicot 1964:204–5) have not been fully answered.

DONN BAYARD
Department of Anthropology, University of Otago.
Box 56, Dunedin, New Zealand. 11 VIII 89

My research interests are limited to the linguistic and archaeological prehistory of Southeast Asia and Oceania over the past 7,000 years or so and the sociolinguistics of contemporary New Zealand. Hence I cannot comment on Bateman et al.'s detailed genetic evidence and such concepts as "topodemes," but my own experience certainly leads me to agree with their refutation of Cavalli-Sforza et al.'s grand scheme.

1. I have been suspicious of phenograms as "God's truth" since cranial analyses from my excavations at Non Nok Tha, Thailand, yielded completely different affiliative dendrograms depending on whether metrical or non-metrical variables were employed (Pietrusewsky in Bayard and Solheim n.d.). As I commented on a similar study of Southeast Asian and Oceanic cranial morphology, "It is clear that small variations in sample size or variable selection can produce vastly different clusters and hierarchies" (Bayard 1987:115).

2. As Bateman et al. point out, only a tiny minority of the world's linguists would accept the validity of entities like Nostratic, Eurasiac, or even Austroasiatic (or even Japanese? [Reid 1988:32]), it is difficult to take such larger classificatory entities seriously.

3. Given their criticisms, Bateman et al. seem strangely hopeful about the rationale for such a scheme in general but call for the gathering of substantial bodies of data from geographically localised populations followed by "multivariate ordination." Sociolinguists have been doing this for some 25 years now. The results indicate the extreme complexity of linguistic interaction and transmission among "extant hominids" [why not "people"?] and make it clear that any tidy division of language transmission into vertical, oblique, and horizontal "components" is a gross oversimplification.
In short, I am a bit puzzled by the “torchbearer” conclusion to Bateman et al.'s paper; surely a brief rebuttal would have sufficed. Perhaps I am being overly negative or blinkered, but I would suggest that Bateman et al. douse the torch and concentrate on more realistic and attainable research goals, say, the affiliation of Japanese or the Austro-Tai hypothesis. As they conclude, any global solutions remain a very distant goal indeed.

BEN G. BLOUNT
Department of Anthropology and Linguistics,
University of Georgia, Athens, Ga. 30602, U.S.A.
16 viii 89

Bateman et al. make a compelling case that reconciliation of human phylogeny and the history of language is premature. As they point out, the lack of a “linguistic clock” renders even the superimposition of language phylogeny on population (genodeme) phylogeny highly problematic. There are no clearly established units of language that can be used to measure change, at least not to any appreciable time depth. As they further note, languages, unlike genes, are not constrained to a reproductive cycle or preprogrammed for replication. Biological evolution and language evolution are thus substantially different in these terms.

One aspect of the difference merits further discussion. Since languages are not directly constrained by the genome, they can undergo radical and extensive change in a shorter time span than the genodeme. Assuming that pressure for language change is not a rare phenomenon, a high degree of variation within and diversity across languages (although with mitigating factors such as borrowing and convergence) may be expected. Reconstruction of language protoforms thus will be more accurate to the extent that variation is documented.

Linguists have long been sensitive to the need for historical documentation, but within a particular perspective. The prototypical form of language reconstruction, as Bateman et al. note, is a description of structural change from time 1 to time 2. Historical information allows for more, and shallower, time points and thus an enriched record of the increments of change. Variation has been viewed by historical linguists principally as either finer gradations on a phyletic branch or points of origin for new branches; beyond this its importance has been neglected. It has not been viewed in the way that biologists typically treat it, i.e., as an integral part of the evolutionary process, in which selection acts on the extent variation in a genodeme. The differences in the ways in which biologists and linguists typically view variation cut to the heart of Bateman et al.'s concluding prescriptions.

Linguists typically work with languages as ideal types—not unmindful of the existence of variation but recognizing that the notions of ideal speakers, languages, and communities remove the clutter of social and historical factors from language. Major theoretical positions in linguistics from Bloomfield to Martinet to Chomsky have adopted this stance. While understanding of language has undoubtedly advanced, it has left linguistics relatively ill-equipped to grapple with variation and thus with vital issues in the reconstruction of phylogeny.

Research methods in behavioral science that directly address variation are rarely taught in the linguistics curriculum, and statistics courses are conspicuous by their absence. Even in research that includes variation, as in sociolinguistics, statistics are often absent or poorly understood (see Wardhaugh 1986). Textbooks on statistics in linguistics have, however, begun to appear (Butler 1985, Woods, Fletcher, and Hughes 1986). Linguists interested in reconciling human phylogeny and the history of language should not only converse with biologists but expand their research methodology to include behavioral science. This will prepare them to carry out the subsequent recommendations, which require quantification. It will also increase understanding of biological transmission (per the introductory quotation).

Dialogue between biologists and linguists is not, however, unidirectional. Bateman et al. acknowledge that the success of a language (survival and spread) is due more to the social influence of its advocates than to its genealogy. Sociolinguists would agree, but they would add that one of the most robust findings of their inquiry is that language is a property of social groups and that the success of a language must be viewed in those terms. Language is an expression of group membership and varies as an individual invokes membership in different groups. In a sense, the engine that drives language maintenance and change and thus produces variation is the way in which groups identify aspects of language and enforce them normatively. Language is thus an attribute not only of individuals but of groups. Language variation across gender and across ethnic, class, occupational, and religious groups is enforced and reinforced to the extent that individuals commit themselves to the norms of those groups. The operational taxonomic units in recommendation 8 will have to be individual humans but as members of groups.

Concern about language as an attribute of groups would require that the temporal progression of initial gene divergence → racial divergence → linguistic divergence include social divergence prior to linguistic divergence. The concept of the differentiated social group (sociodeme) as the transitional unit in language change is consistent with Bateman et al.’s recommendation of cladistics as historical algorithms [6], since changes will not be at a constant rate, and it is also consistent with the key recommendation [3] that human populations and languages be delimited by large-scale demographic census and subsequent multivariate ordination to permit the detection of meaningful boundaries of entities.

CATHARINE A. CALLAGHAN
Department of Linguistics, Ohio State University, Columbus, Ohio 43210, U.S.A. 9 viii 89

Bateman et al. are to be congratulated for their cogent multidisciplinary response to the attempts by Cavalli-
Sforza et al. (1988, 1989) to correlate possible population phylogeny with the deep genetic groupings proposed by Ruhlen (1987) and Greenberg (1987). They have performed a special service by delineating the weaknesses, errors, and inconsistencies underlying such amalgamations as "Amerind."

In this respect, some readers may be unaware of the vast differences among Amerindian languages. My own research has focused on the Utian [Miwok-Costanoan] languages of Central California. I would estimate that Southern Sierra Miwok and Mutsun Costanoan are roughly as far apart as early Latin and Sanskrit. Through the discovery of regular and recurrent sound correspondences involving stems and affixes, I have been able to prove Utian to be a genetic unity, and I have reconstructed a fair portion of the protolanguage. When we compare Proto-Utian with Yokuts [another "California Penutian" family of languages], genetic kinship is already controversial despite some striking similarities. The nearby Pomo languages are utterly unlike Utian, and this without our even having left Central California.

Greenberg disdains comparison based on sound correspondences because they will disappear after a given period. Yet they allow us to identify Southern Sierra Miwok kil-a- 'liver' and Mutsun Costanoan sire 'liver' as perfect cognates even though they do not share a single phoneme. Reconstruction allows us to weed out spurious resemblances such as Sioux ni- and Cheyenne ne- 'second person nominal prefix'. Greenberg may well be right that erosion will ultimately remove the evidence of regular sound correspondences, but erosion will even more quickly invalidate the vaguely perceived similarities of mass comparison.

One of the principal contributions of this article is a systematic discussion of the dissimilarities between the phylogeny of organisms and the evolution of languages. We know that cats are distantly related to catfish because both groups possess backbones and the other structural characters of vertebrates. Languages apparently lack such nonuniversal markers of remote genetic relationship. [As an aside, I must admit that the Utian languages are characterized by metathesis as a morphological process, imperative verbal inflection, and suppletive imperative verbal forms, none of which is a universal attribute of language.] The traits that define linguistic subgroupings are subject to erosion through leveling, replacement, and phonological change. Languages may also borrow extensively from unrelated languages, while exchange of genes between distantly related species is rare [although gene exchange is common between the human populations cited by Cavalli-Sforza et al.].

Given the susceptibility of the nonuniversal features of language to erosion through sound change and other factors, we must question whether it makes any sense to postulate genetic relationship at very remote time depths. After several thousand years, it may well be that all or most of the nonuniversal attributes of the parent language will have disappeared from all or all but one of the daughter languages. At this point, any nonuniversal similarities between the daughter languages will have a nongenetic explanation. The daughter languages will have become unrelated (or "disrelated"), and it will be impossible to recover their common history.

L. L. Cavalli-Sforza, A. Piazza, P. Menozzi, and J. Mountain
Department of Genetics, Stanford University, Stanford, Calif. 94305, U.S.A. (Cavalli-Sforza, Mountain)/Dipartimento di Genetica, Università di Torino, Torino, Italy (Piazza)/Istituto di Eologia, Università di Parma, Parma, Italy (Menozzi).

In this article and in their earlier letter to Science [O'Grady et al. 1989], these critics express their reservations concerning our database, methods of analysis, and interpretations. In the following comments we counter their concerns and point out that their suggestions are unrealistic and unnecessary. Indeed, many of their criticisms surprise us; for example, they have accused us of using gene frequencies based on "very small samples, sometimes of single individuals" (O'Grady et al. 1989). Presumably they are unfamiliar with the data, methods, and models of human population genetics. In fact, apart from one linguist [whose negative comments we know will be answered by Greenberg], they are biologists whose expertise seems to be in taxonomy and the systematics of various zoological and botanical species. Concepts and methods for the study of interspecific differences are not necessarily suitable for that of intraspecific ones, especially in a young species such as modern humans. Because space is limited, we focus on the criticisms we consider most relevant.

Data. Bateman et al. state that our data "do not inspire confidence." Human population geneticists are familiar with the large body of data we use. Tabulations from three books [Mourant, Kopec, and Domaniewska-Sobczak 1976, Steinberg and Cook 1981, Tills, Kopec, and Tills 1983] and 2,815 papers, in part overlapping with those abstracted in the books above, were computerized. From these data on ca. 3,000 "populations" [a statistical and genetic term] identified by the authors of the original publications, geographic gene maps and hundreds of distributions of gene frequencies and sample sizes were generated. These will be presented in a book now in the final stages of preparation. For multivariate analysis, the 3,000 populations were reduced to 42 by eliminating populations and genes too poorly represented and by condensing populations that share an ethnic [or, in a few indicated cases, a linguistic] name and geographic area. We may never obtain the "perfect" body of data that Bateman et al. call for, but the existing collection of data on blood groups, protein and enzyme polymorphisms, etc., which is extremely rich, should certainly be analyzed thoroughly.

Definition of populations. We do not confuse "toplemases," "glottodemes," "genodemes," and "gamodemem." The last two concepts are impractical to use with human populations; the definition of a Mendelian
population, the evidence of mating patterns, and the nearly continuous nature of spatial variation (Cavalli-Sforza and Bodmer 1971) make it impossible to establish nonarbitrary boundaries. Bateman et al. wrongly deny importance to the correlation between geographic and genetic distance (e.g., Morton 1982, Cavalli-Sforza 1984). Barbujani, Oden, and Sokal (1989) have recently developed a strategy to search for boundaries of gene frequencies; its value for building trees remains to be tested.

Missing data. Even though we aimed to maximize the number of genes per population, 23.7% were missing from our final data set of 42 populations and 120 genes. Bateman et al. state that we violate a "general theoretical principle" for which they refer to Temple (1982), but he merely suggests striking out all incomplete rows and columns until one reaches an acceptable percentage of gaps, "say, 5%." This is a suggestion, not a general theoretical principle. By following it with our data set one would lose too much information. Instead, we have treated the problem innovatively with the bootstrap (Efron 1982, Felsenstein 1985). By omitting genes randomly with replacement one estimates the extent of statistical variation due to gene sampling, including that due to missing data. The bootstrap fills a very significant gap in the analysis of trees; we have used the approach to evaluate the compactness of clusters and, with it, the validity of tree nodes. The method of comparing the frequency of clusters observed in bootstraps with that expected if there were no clustering is derived from standard application of the binomial distribution.

Partitions. Bateman et al. consider a partition into six clusters "methodologically incorrect." We named ten clusters simply for expository purposes. They recommend that partitions be based on transverse dissection of the tree at specific genetic distances. This rule assumes unrealistic accuracy of genetic distances and constant evolutionary rates. Moreover, we have repeatedly said that we are not interested in taxometric classifications or their racial implications (see also Cavalli-Sforza and Edwards 1964, 1967; Edwards and Cavalli-Sforza 1964). They further criticize "a posteriori" pooling of five European populations which show very little genetic difference, but their inference regarding inter- and intrapopulation variation is invalid. They do not consider random sampling variation (Weir and Cockerham 1984), the effects of the degree of subdivision on variation of gene frequencies (Cavalli-Sforza and Feldman 1989), or the difference between individual and population variation (Nei and Roychoudhury 1974, Lewontin 1972). Their comparison with mitochondrial-DNA data is invalid for these reasons and because of the different system of inheritance.

Tree reconstruction. Bateman et al. rely on most-parsimonious methods of tree reconstruction only. For metric traits such as gene frequencies, simulations of a simple evolutionary model (Astolfi, Kidd, and Cavalli-Sforza 1981) have shown that a most-parsimonious method does not fare as well as maximum-likelihood or average-linkage methods. No clear genetic model with testable predictions supports the application of maximum parsimony to gene frequencies. We too distrust the uncritical assumption of constant evolutionary rates. We have shown, however, that at least for the major fissions genetic distance is proportional to the time since separation. Our method of tree reconstruction therefore provides a cladistic answer. Methods for species phylogenesis are not always directly applicable to intraspecific data; for example, our critics will find it hard to apply the search for synapomorphies to gene frequencies. They also confuse population trees with gene trees such as those built for the mutational history of mitochondrial DNA. Methods and conclusions applicable to the former do not necessarily apply to the latter (Nei 1987). Finally, we are accused of overlooking anthropometric data; we have omitted them because, unlike genes, they are overly sensitive to selective factors such as climate (Matessi, Gluckman, and Cavalli-Sforza 1979, Piazza, Menozzi, and Cavalli-Sforza 1981). Neutral genetic markers are best for reconstructing evolutionary history (Cavalli-Sforza and Edwards 1967).

Congruence of genetic and linguistic trees. Three things struck us in the comparison of the genetic tree and the linguistic classification: (1) Almost all the linguistic phyla (= families) correspond to low-level genetic clusters or single populations, indicating [a] that the evolution of phyla was relatively late and [b] that successive language replacements were insufficient to blur the general pattern. (2) The major exceptions to this rule can be satisfactorily explained [see below]. (3) The few proposed linguistic superfamilies show correspondence with higher-level genetic clusters.

Consistency index. Bateman et al. compare the genetic and linguistic trees with the consistency index, ordinarily used for measuring optimality. They obtain a value of 0.48, which they declare low, and state that "over half of the linguistic data do not fit." The following errors contribute to this evaluation: [1] They choose to keep the consistency index (as well as their measure of congruence) low, e.g., by rejecting "autoapomorphies" (Brooks, O'Grady, and Wiley 1986), because "phyla confined to a single population are trivially consistent with any tree." A trait present or absent in only one population may be uninformative for tree reconstruction, but in this application they reject strong evidence in favor of congruence of linguistic and genetic evolution. The cases they reject are actually many populations included under a single label. For example, "Australian" includes many tribes, all of which speak Australian languages. When these non-"autoapomorphies" are included, the six "glottodemes" (populations defined by linguistic affinity) still being omitted, the index rises to 0.56. Bateman et al. also keep the value low by considering one language replacement as two events. After correction for this, the index increases further to 0.74 (the maximum being 1.0). [2] Even the 0.48 value they calculate indicates that the two trees are highly consistent. Archie (1989) has shown that the consistency index is negatively correlated with the number of taxa (populations). For a sample of 28 maximum parsimony trees based on real data, the index decreases...
from ca. 0.8 for 10 taxa to ca. 0.2 for 40 taxa. The latter case corresponds most closely to ours [42 taxa]. As 0.48 is much higher than 0.2, Bateman et al. have inadvertently shown that there is indeed remarkable agreement between the genetic and the linguistic tree. [3] They reject a significance test of the index with the spurious argument that "the tree . . . represents a sample size of one." One could, however, generate random permutations of the languages and calculate the consistency index for each permutation. In this way one could obtain the sampling distribution of the index, given the null hypothesis of zero consistency, and proceed to a significance test of the observed index.

Nostratic and Eurasian superphyla. The Nostratic and Eurasian superphyla are recent classifications by independent researchers. Therefore it is not surprising that they differ. They are likely to be modified and extended in the future, and it is reasonable to assume that fully congruent hierarchies will eventually arise. We stated that there is remarkable agreement between the genetic tree and the union of the two sets formed by Nostratic and Eurasian, to which later work has added Amerind languages. The counter-evaluations of our critics do not address this point.

Rotations of tree nodes. Nodes were not rotated "to achieve maximum apparent congruence." The aim was instead just the opposite, to emphasize discrepancies such as the association of Ethiopians with Caucasian speakers of Afroasiatic languages. When this was not technically feasible, e.g., for Tibetans, we had to highlight the exception with a star or discuss it in the text.

Discrepancies between the two trees. We listed and discussed all the discrepancies between the two trees, postulating that they were due to language replacement, a historically well-known phenomenon. In some instances it may be difficult to distinguish this explanation from the complementary hypothesis of gene replacement. The anomalies represented by Ethiopians and Lapps are far from being inconveniences: both populations are mixtures of two genetic clusters, belonging linguistically to one and genetically more to the other. The hypothesis that when two populations speaking different languages mix only one language will survive is certainly plausible. The most likely scenario for the process is gene flow from a neighboring population. This has been observed historically [e.g., for Black Americans] and can easily generate the observed proportions of genetic admixture in realistic periods of time. Admixture evaluations are standard [Cavalli-Sforza and Bodmer 1971], and the method used is described by Wijsman [1984].

Points of agreement with Bateman et al. We have expressed ideas similar to those of Bateman et al. on the Basques [Cavalli-Sforza 1988, Piazza et al. 1988]. We join several other participants in this discussion in pleading for advanced quantitative research in evolutionary linguistics. We agree with these critics on the importance of reciprocal education of biologists and linguists. Two of us have promoted workshops specifically for this purpose, at Stanford in 1984 [organized by L. L. Cavalli-Sforza, M. Feldman, and W. S-Y. Wang] and at Turin in 1988 (organized by L. L. Cavalli-Sforza, A. Piazza, P. Ramat, and W. S-Y. Wang).

The central question is why there should be any congruence between genetic and linguistic evolution. The main reason is that the two evolutions follow in principle the same history, namely, sequence of fissions. Two populations that have separated begin a process of differentiation of both genes and languages. These processes need not have exactly constant evolutionary rates, but rough proportionality with time is a reasonable expectation for both. They should therefore be qualitatively congruent, except for later events such as gene flow or language replacement. These may blur the genetic and the linguistic picture, but our conclusion is that they do not obscure it entirely.

Bateman et al. believe that future research at the regional, rather than global, level may tell us if there is any correlation between linguistic and genetic evolutions. Such research already exists [see, e.g., Spuhler [1979] on North Amerinds, Greenberg, Turner, and Zegura [1986] on the Americas, and Sokal [1988], which lists other examples, on Europe]. Two very detailed unpublished investigations also give highly significant correlations. One, on Sardinia [Piazza et al. 1989], shows a correlation of 0.80 between genetic and linguistic data; the other [Barrantes et al. 1989] gives a correlation of 0.74 for Chibcha-speaking groups of Panama and Costa Rica. Our analysis demonstrates that the two evolutions are remarkably congruent at the global level and generates results of greater time depth than is possible with regional studies.

JOSEPH H. GREENBERG
Department of Anthropology, Stanford University, Stanford, Calif. 94305, U.S.A. 22 VIII 89

My comments are concerned with the rejection by Bateman et al. of much of the linguistic classification found in Cavalli-Sforza et al. [1988]. The agreement with the genetic data will not be within the scope of these comments, although others have found it impressive.

Bateman et al. perceive such difficulties in distinguishing historically significant resemblances from accidental convergences that it becomes difficult to see how linguistic classification can get started. "Perceived linguistic similarities (whether they reflect phonology, grammar, semantics, or syntax) have a highly problematic relation to historical connections." In this context terms such as "phonology" are far too vague. They disregard a fundamental property of language, namely, the in-principle arbitrariness of the relation of sound to meaning in specific forms, whether lexical or grammatical. Moreover, languages have large numbers of such signs which are essentially independent of each other. It makes sense to distinguish, as typological and without necessary historical significance, resemblances in sound only [e.g., tonal systems in Southeast Asia and in Africa] and in meaning only [e.g., gender in Chinook and in French] from resemblances involving sound and meaning simultaneously [e.g., German gut, English "good"]
and Hebrew -t and Arabic -t, both marks of the feminine gender. It was the elimination of typological criteria that was the single most important factor in the success of my African classification. The vast majority of convergences are typological. Those involving sound-meaning similarities must be diligently sought. Far from being handicapped in relation to biological classification, linguistics has great advantages because of the basically arbitrary relation between form and function and the number of essentially independent form-meaning sets.

The concrete example that Bateman et al. adduce is second person singular Sioux ni, Cheyenne ne, and Ojibwa ke, in which the Algonquian languages Cheyenne and Ojibwa differ while Cheyenne agrees with Sioux. Such instances fool no one. In Greenberg (1987a) we find Algonquian k, second person, and I did not connect the Sioux and Cheyenne forms. How did they detect it? They used the comparative method, but this implies the prior classification of the Algonquian languages as a group within which k is obviously normal and n deviant. According to them, linguistics cannot classify until it has distinguished true resemblances from convergences, but, as we have just seen, this can only be done on the basis of a classification which has already been carried out. We are caught in a vicious circle.

Just how frequent are accidental form-meaning convergences in language? A whiff of reality dispels this supposed difficulty and reveals a truly gargantuan gap between their theory and reality. In Greenberg (1988:24) a table of nine basic words in 24 European languages is presented. Before one has gotten to the third word the basic division into Indo-European, Finno-Ugric, and Basque appears, and by the fourth word one has the accepted subgroupings of Indo-European. All the similarities immediately perceivable are incorporated in standard Indo-European etymological dictionaries, and some that are less obvious are to be found. This is the result of almost two centuries of investigation by a substantial portion of the linguistic community. One cannot reject all these obvious similarities as illusions without destroying comparative linguistics.

From the qualification "s.l." [sensu lato] after "Na-Dene" it appears that the authors must still doubt the affiliation of Haida despite Greenberg (1987a:chap. 6, cf. also 1987b:651–52). Levine (1979), a supposed refutation of Sapir, has been cited by Goddard (1986:191) as the kind of critique of Sapir's work that is needed. As I have shown, even after applying Levine's largely unreasonable criteria 17 etymologies remain linking Haida, Tlingit, and Athabaskan and 14 linking Haida and Athabaskan. The same criteria applied to three branches of Indo-European leave only 6 by a liberal application of Levine's rules, probably none if they are applied strictly. This shows that, if anything, Na-Dene as proposed by Sapir is a more tightly knit group than Indo-European and that the methods approved by Goddard would refute even Indo-European. It also shows that the family-phyllum distinction the authors wish to maintain, by which Indo-European is a family while Na-Dene is a phylum, is a property not of languages but of linguists. The abolition of the term "phyllum" would be a methodological advance, since it merely adds a rhetorical notion of remoteness and suggests that a different and more mysterious methodology is involved.

Finally, the statement "These difficulties may not be insuperable, particularly if initial studies sample a regional rather than a global catchment" is but one more instance of the strange belief that the narrower one's factual basis, the more successful one is likely to be. The enormously widespread n first person and m second person in the Americas as against m first person and t second person in Europe and Northern Asia is powerful evidence which requires an explanation. Bateman et al. have neither the capability for making such observations nor the type of theory which can account for them.

KENNETH JACOBS
Département d'Anthropologie, Université de Montréal, C.P. 6128, succursale A, Montréal, Qué., Canada H3C 3J7. 4 VIII 89

This article, neither fish nor fowl, leaves me somewhat confused as to its purpose. The authors attempt both to critique an article by Cavalli-Sforza et al. and to present their own methodological scenario for the establishment of patterns of congruence between linguistic and other aspects of human evolution ("human" being a term inadequately defined). They have clearly underscored an important issue: to what extent material-cultural [archaeological], linguistic, and human morphological data can be employed to reconstruct the genealogy [phylogeny seems optimistic] of recent human populations. That the data set used by Cavalli-Sforza et al. is ill-chosen or the statistical analyses inappropriate may be true, I have not the training to judge.

One aspect that troubles me, however, is the introduction into the text of neologisms such as "glottodeme," "topodeme," "entyy," and so on, without, to my mind, adequate definition. Especially in a journal that reaches out to all kinds of anthropologists, a simple reference to the article in which the term was coined seems, at best, insufficient. A footnote defining terms might have aided comprehension.

As a biological anthropologist I am especially troubled by the irregular and ambivalent use of the terms "race" and "racial." At times, these terms appear in quotes, implying a certain disdain for the concept that there are biological races. At times, "racial entities" are accepted as real and scientifically definable. This latter point of view is reinforced when we read that "several techniques allow objective delimitation of genetic races." While perhaps being a useful means for the authors to divide their own data base, "race as such," as Harrison (1988:326) has commented, "explains practically nothing." Biological anthropologists having increasingly given up the racial paradigm, a fuller explanation of how we can set the standards for an "objective" linguistic cladogram would have been welcome.

The authors are to be commended for their efforts. No entente cordiale should be expected, however, since the
modes of transmission of material and linguistic culture, much less those of the genome, are only cursorily and confusedly discussed. Without more detailed discussion, the very tempting and important questions they broach will remain unanswered. Still, they have made a first, if halting, step.

YUII MIZOGUCHI
Department of Anthropology, National Science Museum, 3-23-1 Hyakumin-cho, Shinjuki-ku, Tokyo 169, Japan. 17 VIII 89

Most of the arguments and conclusions of Bateman et al. seem reasonable. Among the problems they point out, however, that of developing objective techniques for quantifying language change has already been partly resolved by mathematical linguistic methods such as Os- walt's shift test (Yasumoto 1978). Increased discourse between linguists and biologists will be more easily achieved in the near future, but, as the authors themselves realize, it is difficult to apply specifically phyl- ogenetic algorithms such as cladistic ones to data, whether genetic or linguistic, on modern populations only because of difficulties in distinguishing derived characters from primitive ones. To trace the evolution- ary process of appearance, retention, disappearance, or reappearance of any character, fossil or past evidence is needed. Although DNA investigations of fossil samples have recently become possible to some degree, it will take a long time to accumulate such data on a scale comparable to that of morphological data on fossils.

Problems concerning the delimitation of entities or populations, the definition of characters or attributes, and large-scale sampling are much more fundamental and in practice very difficult to resolve. At an early stage of investiga- tion, therefore, preliminary analyses such as those performed by Cavalli-Sforza et al. (1988), even if based on insufficient data, should be allowed as a stepping stone.

At present, if changes in attributes such as DNA in the course of evolution are regarded as stochastic, the most likely of possible phyllogenetic trees can be deter- mined by Felsenstein’s maximum-likelihood method without any assumption that the rate of evolution is constant in all the organisms studied (Hasegawa 1988), and the statistical significance of the maximum loga- rithmic likelihood can be tested by Efron’s (1982) bootstrap method (Hasegawa 1988). But linguistic and morphological attributes and even some genetic ones are not simple stochastic variables. The construction of a phylo- genetic tree on the basis of these attributes therefore requires a different procedure. Whether the procedure used is cladistic or phenetic, the distinction of primitive and derived characters and their susceptibility to environ- mental factors is very important in interpreting the results. The effect of environmental factors may be esti- mated to some degree for some characters, but, as men- tioned above, the derivation of characters cannot be de- termined without historical evidence. Therefore the best procedure for reconstructing phyllogenetic lines seems to be to compare data sets from two groups of populations at two different but contiguous times using phenetic methods and to repeat the comparison for each past time span. This is possible to some extent for morphological and linguistic data and may soon be so for DNA data. Linguistic inheritance is different from genetic inheri- tance in some ways, to be sure, but linguistic data still seem also useful for reconstructing the microevolutionary processes and migrations of human populations dur- ing historic times provided that we know the character- istics of the linguistic attributes. Once a series of such two-group comparisons has been carried out, the cladi- stic procedure will of course also be available.

Incidentally, the overall correlation between a biological tree and a linguistic classification may be estimated and statistically tested by Mantel’s matrix permutation inference procedure (Mantel 1967, Dietz 1983, Dow and Cheverud 1985) if the distance matrices based on such data are phenetically estimated. Furthermore, canonical correlations between biological and linguistic attributes may be estimated on the basis of ecological correlations obtained by the use of mean values or some other statistic on each sample, and their significance can be tested by the above-mentioned bootstrap method.

In any case, Bateman et al.’s paper is very stimulating. I share their hope that a more accurate human phyllogenetic tree can be developed through the integration of genetic [both morphological and biochemical] and linguis- tic data.

MILTON NUNEZ
Bureau of Museums, SF-22100 Mariehamn, Åland Islands, Finland. 27 VIII 89

Bateman et al. have succeeded in presenting a clear, thorough discussion of the problems faced in attempting to reconcile the evolutionary history of humans with that of their languages. Unfortunately, the omission of key information by Cavalli-Sforza et al. (1988) hinders not only the appraisal of their results but also the ade- quate evaluation of Bateman et al.’s serious objections. I refer to questions such as whether one can allow 23.7% gaps in a data-matrix, even for the noble purpose of max- imizing attributes. The objections of Bateman et al. to such unorthodox procedures certainly seem justified. The same applies to the “population means” used by Cavalli-Sforza et al.; these critics are probably right in predicting greater variation within single populations than between different “population means.” I suspect this to be the case at least with the Lapps.

It should be pointed out, however, that Cavalli-Sforza et al. (1988) appear to place great confidence in the ability of the “bootstrap” method to test the validity of their data. I am not qualified to judge; nor do Bateman et al. comment on this new and seemingly useful statistical tool. The core of the argument lies in whether rigour re-quirements may be relaxed if results are subsequently submitted to “bootstrap” testing. Cavalli-Sforza et al. think so; Bateman et al. do not.
The use of recent, albeit controversial, linguistic phyla by Cavalli-Sforza et al. [1988] appears justified in that the lumping approach of these classifications better fits the genetic tree. Correspondence with the tree would not have significantly suffered by using the more accepted classifications, at least in the case of certain controversial phyla [e.g., Altaic, Amerind, Na-Dene]. Their use has basically a cosmetic effect on fig. 1. We must also bear in mind that the controversial phyla and superphyla were derived independently by others and, consequently, their correspondence with the genetic tree cannot be dismissed without explanation. Furthermore, widely accepted phyla result from traditional methods of historical linguistics, which fall out of range beyond a few thousand years B.P. Despite claims of shallow time-depth by Bateman et al., “mass comparison” methods lack the limitations of traditional historical linguistics and may therefore be better equipped to work at greater time-depths. Undoubtedly a better alternative, endorsed by Bateman et al., would be the quantitative approach suggested by Diamond [1988].

Bateman et al. have shown that the correlation between human and linguistic evolution may not be as straightforward as the conclusions of Cavalli-Sforza et al. would suggest. Nevertheless, I cannot help feeling that they come down too hard on the pioneer efforts of Cavalli-Sforza et al. and that perhaps we should withhold judgement until more is known about their methods and data. We are all aware of the differences in the evolutionary processes affecting living organisms and languages, and it is precisely this that raises doubts about their possible correspondence. On the other hand, for 50–100 millennia language-bearing human groups have been chronically interacting with their neighbours. During periods of separation both their genes and their languages have evolved differentially in isolation, during contacts the exchange of genetic and linguistic elements has taken place to varying degrees and in varying ways along with other aspects of culture. In other words, despite the fact that they are subjected to somewhat different processes and rates of transmission, adoption, and preservation, both genetic and linguistic elements tend to be passed on and received during the same interaction periods. It is perhaps traces of these very processes that we see in Cavalli-Sforza et al.’s results.

A more reliable picture may be obtained by following Bateman et al.’s recommendations, but I personally would like to see archaeology playing a more important role. Culture, including language, is often transmitted together with genes, and therefore archaeological data may serve as a useful control.

ROBERT L. OSWALT
California Indian Language Center, 99 Purdue Ave., Kensington, Calif. 94708, U.S.A. 29 VIII 89

These comments are on both the critique [Bateman et al.] and the critiqued [Cavalli-Sforza et al. 1988]. I found the critique the tougher reading, reduction of the amount of specialized terminology, as a concession to the general reader, would have been appreciated.

Figure 1 of Cavalli-Sforza et al. [reproduced as figure 1 here] shows a coincidence of tree diagrams based on genetic and linguistic factors. Whether the degree is remarkable depends on expectations. Some parallels and some divergence are to be expected, for reasons discussed in Bateman et al., who also point out many of the inadequacies in the statistical treatment underlying the genetic tree and the controversial nature of the linguistic classification at the superphylum level. I agree with their criticisms but, rather than speaking in general methodological terms, will illustrate the development of my opinion on the subject through an examination of concrete results at the more local level.

The first anomaly to come to my notice is the one linking “Uralic (Ling.)” closely with Mongol and only after six nodes with the Caucasoid group. Uralic is a common linguistic term for a stock as ancient and well-established as Indo-European [see Anttila 1972]. Cavalli-Sforza et al. have taken a sample of “Samoyed and other Uralic language speakers living near the Ural mountains” and labeled it “Uralic (Ling.).” They apparently believe that their sample is representative of Uralic as a whole, as they later state, “Lapps associate linguistically with speakers of Uralic but genetically with Caucasoids.” The anomaly is not of the Lapps, who are simply Caucasian along with millions of other Caucasian speakers of the western Uralic languages, but of the two branches of Uralic: the divergent, small (in population) eastern (Samoyed) linguistic branch apparently fitting better in the Northeast Asia genetic group and the overwhelmingly more populous western (Finno-Ugric) branch being Caucasian. At the superphylum level the evidence is fairly strong that Uralic and Indo-European are related [Pedersen 1962 [1931], Oswalt n.d.], supporting the conclusion that it is the eastern branch of Uralic that has shifted its genetic affliation.

Cavalli-Sforza et al. say that by their bootstrap test Berbers leave the Caucasian cluster 20% of the time and tend to join the African cluster. A shift 20% of the time sounds substantial to me but would perhaps be acceptable if Berber were simply insecurely placed between two very recently formed branches. Berber, however, is shifting between the two branches of the first bifurcation of all mankind. With such results, one can wonder about the appropriateness of a tree diagram for displaying the data. Would not some technique for presenting clinal relationships be more appropriate? Cavalli-Sforza et al. argue against one such model of an almost continuous population network [that of Wolpoff, Wu, and Thorne 1984], but I still do not understand their resistance to utilizing something like multidimensional scaling. If the calculations are too formidable, that is just another reason to work out the methods on the local or regional level before going global.

In the intriguing region through which the Americas were peopled, Cavalli-Sforza et al. find the Eskimo closest to the Chukchi and closer to the peoples of Northeast Asia [including the Japanese and Korean] than to North-
west Amerind or any Amerind. This is potentially a truly important piece of evidence unfortunately diminished by their statement that the cluster is "not very tight" and by doubts raised by other results. The precise composition of the sample is critical here: Is the "Northwest Amerind [Na-Dene Ling.]" population made up mainly or entirely of Indians far removed from the Eskimo [Navajo or Haida, perhaps], or is it a nicely spaced sample? To me, the presentation would be more revealing if instead of such large-scale pooling samples had been taken of smaller divisions of the Eskimo-Aleut and of the divergent languages grouped under Na-Dene.

Although it might not be apparent from the recent dates on the references to linguistic classifications, discussion of the relationships between genetic and linguistic groupings has a long history, even as do many of the present proposals for far-reaching linguistic affiliations, especially those in the Eurasian landmass. William Dwight Whitney, in a series of lectures in 1864 and 1865 [Whitney 1872: esp. Lectures 8–11], elegantly and clearly discussed the topics of close and distant linguistic classification, the relationship between race and language, and the relative merits of "lumpers" and "splitters" [not his terms]. We are accumulating more and better data and improving statistical techniques, yet a great deal of what was said over 100 years ago could be usefully consulted now.

[Bateman et al. will respond in the next issue.—Editor.]

References Cited


