On the Pattern and Origin of Human Biological Diversity

Kenneth Korey

Following its turbulent response to Carleton Coon's 1962 publication of *The Origin of Races,* anthropology's interest in race grew increasingly quiescent over the next decades. Only within recent years has race returned to center-stage. The reasons underlying both our withdrawal from racial study and our current reinvestment in it are unquestionably varied: some are ideologically rooted, while others have more empirical foundations. I shall concern myself hear with only this latter set of factors—those ostensibly belonging to the class of empirical discovery. First, I shall consider briefly a signal development in the widespread rejection of the racial paradigm on genetic grounds. Next, I want to consider how reinterpretation of the same class of genetic data has fed a revival of our disciplinary interest in race. In the end, I want to draw the two lines together, in a consideration of whether or not—at least from the genetic perspective—there seem to be justifiable reasons to resurrect race as an organizing schema for human diversity.

The first matter can be disposed of quickly: the pre-eminent empirical factor in our post-Coon detachment from race was Richard Lewontin's (1972) initial quantification of the large-scale patterning of human genetic diversity. Certainly his conclusion that the traditional construction of race served as an inadequate descriptor of human genetic variability reinforced anthropology's retreat from racial thinking, a withdrawal already well under way since the second World War. Can anyone among us be unaware that most human genetic variation is intra-populational, and that only a small fraction of the total separates what conventionally we once regarded as the major human races?

Shifting attention to the present, restoration of the racial paradigm is being empirically driven by renewed attempts to reconstruct a large-scale history of human biological variation. (This is not to suggest that race has become insinuated in *all* current efforts to resolve our evolutionary history, although frequently it is. For example, while Cavalli-Sforza et al. [1994] have claimed to place their work explicitly outside a raciological framework, they persist in designating aggregates of geographically contiguous populations as "Negroid," "Caucasoid," and "Mongoloid.") While this has been our oldest and most enduring project, since mid-century our labors have centered on the interpretation of patterns of genetic, rather than morphological, variation.

The recurrence of the racial construct found in contemporary work does not deny the *facts* about the variational pattern presented by Lewontin; to the contrary, these facts have found only confirmation (e.g., Nei and Roychoudhury 1974; Latter 1980; Ryman et al. 1983; Relethford 1994). Instead, it denies what initially appeared to be Lewontin's unassailable *criterion for utility* in categorizing patterns of diversity: that there out to be more variation between constructed classes than within them.

Lewontin's criterion appeared satisfactory insofar as our approach to the question of race was exclusively grounded in nominalist terms: *Do races categorize patterns of biological variation appropriately?* The answer, of course, was *No*. But the attempt to derive biological histories from these same data introduces an altogether different criterion for utility: *Are races phylogenetically identifiable, historical entities which have produced the now-observable patterns of variation?* Increasingly, the answer for many seems to be *Yes*. It is important to understand that as phylogenetic concerns displace classificatory concerns, the criterion of utility shifts from *nominalist* terms to *realist* terms. This is why it is possible for some to argue that races really do exist (presumably as historic entities) with full knowledge that they express only a minor component of human variation (e.g., Sarich 1995). Utility, then, is conditioned on purpose.

The interpretation of genetic data has often been predicated on the assumption of primordial races, from the earliest ABO blood group data—taken to represent the imposition of a pure A race and a pure B race superimposed on a primordially O human species (Marks 1995). But what if it were shown that patterns of variation often attributed to descent from geographically coherent and identifiable founder stocks—*races*, in this realist view—were instead products of something other than successive splitting and divergence of ancestral stem populations or stocks? Would we not be forced to rethink again the racial paradigm that some workers now allege to represent historical events?

Nei and Roychoudhury (1993:936-7) constructed a phylogenetic tree for 15 representative populations from allele frequencies at 33 nuclear loci. By their account, the tree reveals that "human populations can be subdivided into five major groups: (A) negroid (Africans), (B) caucasoid (Europeans and their related populations), (C) mongoloid (East Asians and Pacific Islanders), (D) Amerindian (including Eskimos), and (E) australoid (Australians and Papuans)." Further, they write that "the evolutionary relationships of these major groups are hierarchical rather than parallel, and some groups apparently originated from a population belonging to some other groups. . . ." (937). Thus they find support for the earliest major split between Africans and non-Africans, the next between Europeans and all other non-Africans, and so forth (Figure 1). The racial identities of the outermost branches obtain, historically, from their putative derivation from regionally differentiating ancestral stocks.

In a more simplified schema representing the Old World populations, these large-branch features can be seen in the equivalent and derivative topology given in Figure 2a. Here the terminal branches are replaced by boxes, whose lengths are the mean branch lengths of the populations aggregated within each region and whose widths are scaled to the average population heterozygosity within each region. Each box thus provides a graphic representation of intra-regional genetic diversity.

The question we may pose is whether it is possible and plausible that such a typology as this could have been produced by some conflation of evolutionary processes other than those which implicitly underpin the succession of population splittings depicted. One way to answer the question is by attempting to build a Monte Carlo simulation to generate the same pattern of current human diversity that informs constructions of this tree, but without incorporating the hierarchical succession of racial subdivision it presupposes.

Imagine, for example, a world subdivided into four regions, each large enough for up to 50 populations of arbitrary size. Imagine that in some initial stage a population, represented by 500 unlinked diploid loci (each with up to three alleles), were to arise at random in some region, and subsequently to expand into adjacent vacant neighborhoods, radiating at a distance of one adjacent tier every generation (Figure 3a). When this region first became fully occupied, imagine the other three regions to be penetrated and similarly colonized (Figure 3b), until all became filled (Figure 3c).

Suppose that all this takes place relatively rapidly, so that at the conclusion of the initialization process—when global colonization is complete—the differentiation among regions is negligible. From here one, all populations remain geographically stationary and the reckoning of time begins. However, for simplification, assume that:

- They undergo mutation at a rate of 5×10^{-6} mutations per locus per generation;

- They undergo drift at every locus, at which selection exerts negligible effects;

- They exchange genes intra-regionally at specified rates according to a two-dimensional stepping-stone migration model; and

- They exchange genes inter-regionally through migration between designated proximal populations at specified rates.

This structure leads to two Monte Carlo models, distinguished by the absence (Model 1) or presence (Model 2) of inter-regional gene flow. In both cases, the

centrifugal tendency of regional populations to disperse genetically, powered by mutation and drift, is a function of the size and number of local populations within each region and the rates of gene exchange among them. On the other hand, the *centripetal* effect of gene flow, where it connects these regions, draws them closer together. These oppositional tendencies can be balanced by varying rates of intra- and inter-regional gene flow and by varying the numbers and sizes of constituent populations. Since these factors jointly determine variance effective numbers in a hierarchically subdivided species, it becomes possible to govern its genetic structure—and, importantly in this context, to regulate the rates and magnitudes of regional divergence—by judicious choice of parameter values.

In even so simple a model as this, which values might produce the desired effects is surprisingly difficult to determine analytically. It was simplest to adopt a recursive approach, building up a large catalogue of outcomes under systematically varied trial values, and to select those parameter sets whose outcomes most closely approximate the required independent variables. Using the regulatory values given in Table 1, I allowed the simulations to run 4000 generations (roughly equivalent to 100,000 years) according to the two models noted above.

Comparing the Nei-Roychoudhury tree to the Model 1 tree (Figure 2b), both were obtained by identically applying the neighbor-joining algorithm (Saitou and Nei 1987) to the genetic distance, D_A (Nei et al. 1983). Both trees were rooted by equal partition of the greatest inter-regional distances. The two typologies are identical in their branching structure, and equally robust (shown by bootstrap resampling) in the determination of their branch patterns. Both show Region 1, representing Africa, to be the outlier—but to be perverse, I allowed the initial founding population to originate in Region 4, Australasia. The intra-regional parameters—average heterozygosity and inter-population distances—do not differ significantly ($p<0.05$) from the values estimated from Nei and Roychoudhury's data (Table 2). Notably, in light of Lewontin's original analysis of human diversity, the apportionment of total species heterozygosity approximates that estimated from the real world: 85% within populations and 10% between regions (Table 3).

The reasonable fit of this simple model to the Nei-Roychoudhury tree strongly implies that the interpretation of the fundamental pattern of human genetic variability as the product of successive subdivisions of ancestral stocks is gratuitous. Similarly, the conclusion of Bowcock et al. (1991), that attachment of the short European branch to the Asian stem indicates greater Asian than African contribution to an ancestral European gene pool derived from early admixture, need not follow—and, in the case of Model 1, cannot possibly follow. Finally, neither does the greater proximity of European and Asian populations to one another relative to African populations (see, for example, Nei and Livshits 1989) correctly support an inference for the African origin of the modern species.

Model 2 shows similar conformity with the Nei-Roychoudhury tree (Figure 2c). Regions 3 and 4 cluster consistently together, and Region 1 is always the outlier (although the real origin was located in Region 4, as before). The intra-regional parameters are again not significantly different between the two trees (Table 2), and the apportionment of total species heterozygosity again approximates 85% within populations and 10% between regions (Table 3).

Again, may we rightly conclude that the pattern of variation is the necessary consequence of a hierarchical subdivision of primordial racial stocks? Hardly! After 4000 generations the most cosmopolitan area in the simulation, Region 4 (representing Australasia), with the highest rate of immigration, will have fewer than one-fourth of its genes originating autochthnously, while the most isolated region, Region 2 (Europe), will have no more than two-thirds. In terms of gene identity, this looks more like an amalgam than a racially subdivided species.

Significantly, valid divergence times cannot be estimated for any of the internal nodes of the tree. Indeed, no such nodes corresponding to singular historical divergence events even exist—merely their virtual images as analytic artifacts.

My conclusion is simple: the patterns of genetic diversity in the human species greatly underdetermine a topology of hierarchical subdivisions intended to represent an underlying history. If a model as absolutely crude as the one I have illustrated—one disallowing long-distance migrations, natural selection, unequal rates of gene flow within regions, variation in population size within regions, and so on—if a model so crude as this can reproduce by other means the fundamental tree-like structure popularly envisioned as the evolution of genetic diversity, then we ought strenuously to question any claim to knowledge of a human racial history. And if our concept of race cannot be situated at the root of our history, can it be usefully situated *anywhere*?

References Cited

Bowcock, A. M., J. R. Kidd, J. L. Mountain, J. M. Hebert, L. Carotenuto, K. K. Kidd, and L. L. Cavalli-Sforza
    1991   Drift, Admixture, and Selection in Human Evolution: A Study with DNA Polymorphisms. Proceedings of the National Academy of Science 88:839-843.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza
    1994   The History and Geography of Human Genes. Princeton: Princeton University Press.

Coon, C. S.
   1962   The Origin of Races. New York: Alfred Knopf.

Latter, B. D. H.
   1980   Genetic Differences Within and Between Populations of the Major Human
           Subgroups. American Naturalist 116:220-237.

Lewontin, R. C.
   1972   The Apportionment of Human Diversity. Evolutionary Biology 6:381-398.

Marks, J.
   1995   Human Biodiversity: Genes, Race, and History. New York: Aldine de
           Gruyter.

Nei, M. and G. Livshits
   1989   Genetic Relationships of Europeans, Asians, and Africans and the Origin
           of Modern Homo sapiens. Human Heredity 39:276-281.

Nei, M. and A. K. Roychoudhury
   1974   Genic Variation Within and Between the Three Major Races of Man,
           Caucasoids, Negroids, and Mongoloids. American Journal of Human
           Genetics 26:421-443.

   1993   Evolutionary Relationships of Human Populations on a Global Scale.
           Molecular Evolutionary Biology.. 10:927-943.

Nei, M. F. Tajima, and Y. Tateno
   1983   Accuracy of Estimated Phylogenetic Trees from Molecular Data. II. Gene
           Frequency Data. Journal of Molecular Evolution 19:153-170.

Relethford, J. H.
   1994   Craniometric Variation Among Modern Human Populations. American
           Journal of Physical Anthropology 95:35-52.

Ryman, N. R. Chakraborty, and M. Nei
   1983   Differences in the Relative Distribution of Human Gene Diversity between
           Electrophoretic and Red and White Cell Antigen Loci. Human Heredity
           33:93-102.

Saitou, N. and M. Nei
   1987   The Neighbor-Joining Method: A New Method for Reconstructing
           Phylogenetic Trees. Molecular Evolutionary Biology 4:406-425.

Sarich, V. M.
   1995   Human Races are Very Real and Very Young. American Journal of
           Physical Anthropology Supplement 20:189.

	Model 1				Model 2			
	region				*region*			
	1	2	3	4	1	2	3	4
population size								
N_e (population)	500	200	600	700	350	250	600	650
No. populations	10	30	20	8	6	18	19	5
N_e (region) x 1000	5	6	12	5.6	2.1	4.5	11.4	3.3
migration between populations								
no. migrants (per generation)	1.25	6.00	3.00	1.25	1.50	7.50	3.00	1.63
immigration from contributing regions								
contributing regions					2,3	1,3	2,4	1,3
no. immigrants (per 10 generations)					6.3	5.1	25.5	34.6

Table 1. Simulation parameter values for Model 1 and Model 2

	region			
	Africa [1]	Europe [2]	Asia [3]	Australasia [4]
Heterozygosity				
Nei and Roy-choudhury (1993)	.274	.307	.278	.201
model 1	.216 *(.203)*	.216 *(.190)*	.247 *(.197)*	.220 *(.206)*
model 2	.194 *(.201)*	.233 *(.189)*	.254 *(.194)*	.223 *(.209)*
Distance				
Nei and Roy-choudhury (1993)	.044	.005	.015	.040
model 1	.038 *(.017)*	.005 *(.002)*	.014 *(.004)*	.042 *(.025)*
model 2	.047 *(.023)*	.005 *(.002)*	.015 *(.005)*	.041 *(.019)*

Table 2. Intraregional diversity, measured by mean intraregional heterozygosities and distances; standard deviations given in parentheses.

Table 3. Apportionment of species gene diversity, standard
deviations given in parentheses

Source	total gene diversity (H_T)	relative distribution, %(s.d)		
		w/in pops	btwn pops w/in region	btwn regions
mixed serological markers (25 loci) [1]	.312 (.055)	86.0 (2.1)	4.1 (0.6)	9.9 (2.2)
model 1 (500 loci)	.283 (.203)	83.4 (11.8)	4.1 (2.1)	12.4 (11.8)
model 2 (500 loci)	.285 (.173)	85.4 (10.3)	4.2 (2.1)	10.4 (10.1)

[1] *Ryman, Chakraborty, and Nei (1983)*
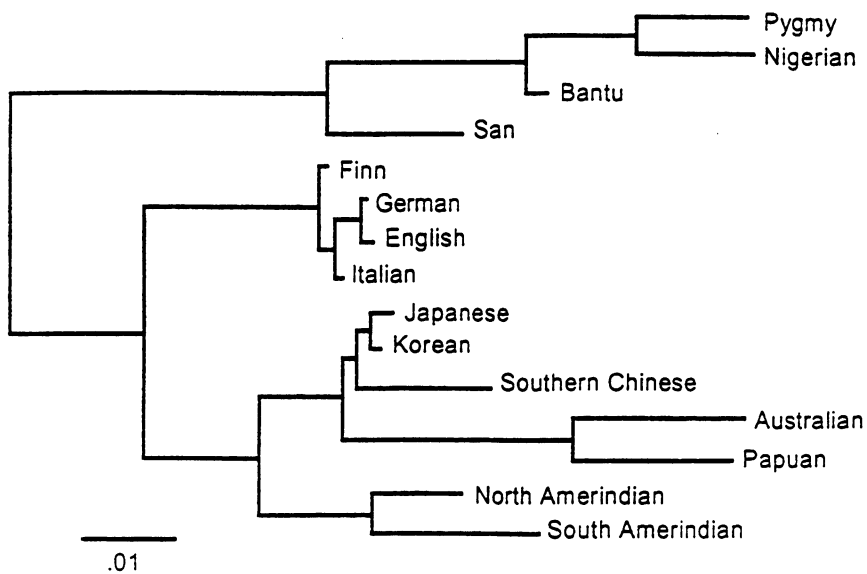
FIGURE 1. Neighbor-joining tree for 15 representative populations,
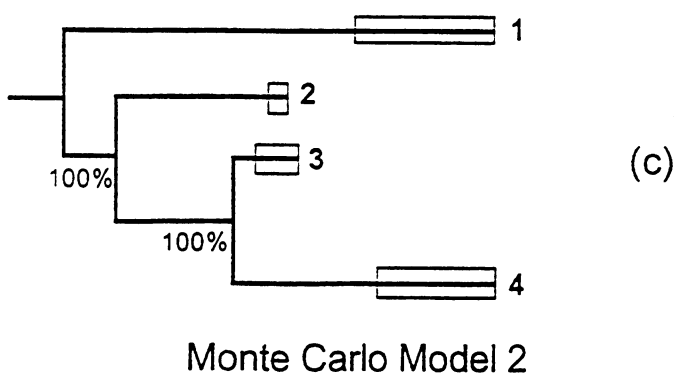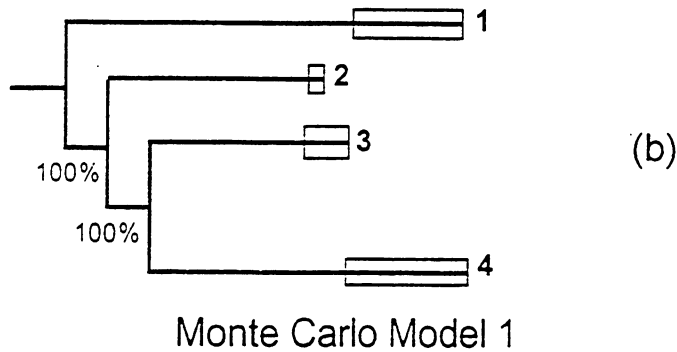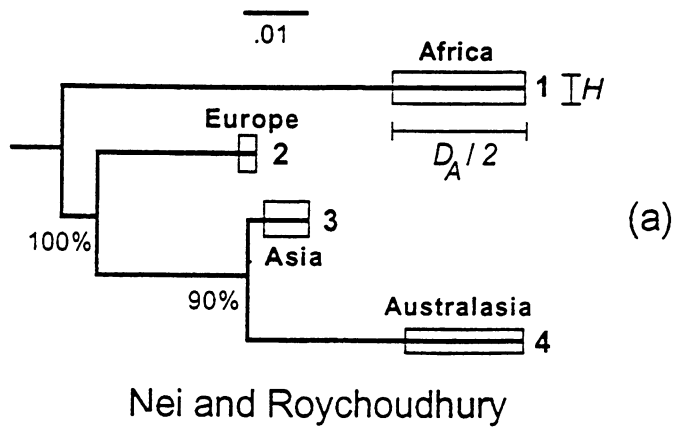redrawn from Nei and Rouchoudhury (1993).

FIGURE 2. Simplified neighbor-joining trees representing four Old World regions depicted in Figure 1, with rectangles representing intra-regional means for heterozygosity, *H*, and distance, $D_A/2$, replacing terminal branches and bootstrap probabilities for 500 random resamplings of loci given as percentages in the two Monte Carlo models; (a) simplification of Nei and Roychoudhury (1993) tree; (b) tree obtained from Model 1, without intra-regional gene flow; (c) tree obtained from Model 2, with intra-regional gene flow.

FIGURE 3. Three stages of Monte Carlo model's initialization.

INITIAL PHASE

(a)

COLONIZATION PHASE

(b)

FINAL PHASE

(c)