5

Race, Ancestry, and Medicine

MARCUS W. FELDMAN AND RICHARD C. LEWONTIN

The cost of assessing DNA-level variation in large numbers of people has steadily declined. There are now large sets of data on variation of different kinds of DNA markers among geographically diverse people. Several technical methods of studying human genetic variation are used, including analysis of single nucleotide polymorphisms (SNPs) (e.g., Voight, Kudaravalli, Wen, & Pritchard, 2006); short tandem repeats (STRs or microsatellites) (Rosenberg et al, 2002; Rosenberg et al., 2005); and Alu sequences (Bamshad et al., 2003). There are now many millions of polymorphic sites revealed in hundreds (and soon, perhaps thousands) of people available for analysis. At the same time, statistical tools for describing and interpreting observed patterns of variation have increased in sophistication.

Recent analyses of hundreds of microsatellite[1] DNA markers and a few thousand SNPs from human populations have shown that it is possible with a high degree of accuracy to assign the major geographical region (or regions) of origin of individual human beings by using a combination of a number of these polymorphic genes (Rosenberg et al., 2002; Rosenberg et al., 2005; Conrad et al., 2006). In addition, using more markers, it is possible in some cases to narrow down the population of origin to local national populations within major geographic regions. The greatly increased facility with which human genetic variation can be studied and the suggestion that this variation may be exploited to individualize medicine have fueled a growing controversy about whether race is indeed a biologically useful and meaningful concept when applied to humans, especially in a medical and pharmacological context. Four commentaries in the *New England Journal of Medicine* (Burchard et al., 2003; Cooper, Kaufman, & Ward, 2003; Phimister, 2003; Wood, 2001), one in *Science* (Sankar & Cho, 2002), an editorial ("Genes, drugs, and

race," 2001) and commentary (Calafell, 2003) in *Nature Genetics*, and four reports in the *New York Times* (Satel, 2002; Wade, 2002a, 2002b, 2003) all raised the issue of the status of racial categorization as a biological concept. In particular it is claimed that these recent data are in contradiction to the widely accepted and confirmed observations that a very large proportion of human genetic diversity lies within geographical regions, observations that have led biologists and anthropologists to abandon the notion of human races over the last 30 years. What we wish to do is to explain that there is no contradiction between these two well-substantiated bodies of data because they speak to two quite different questions that have been confused.

The microsatellite data and data on other DNA polymorphisms (Bamshad et al., 2003) are relevant, among other things, to the problem of the assignment of individuals to lines of geographical ancestry. The question asked is whether it is possible to find genes that are polymorphic in the human species and whose frequencies of alternate alleles are sufficiently different in the different major geographical regions to allow a correct assignment of geographical origin with high probability. The answer to this question is "yes," and that answer has been known for 50 years from studies of genetic polymorphisms. This is a problem in biological systematics.

The data on general genetic polymorphism for proteins and nucleotide substitutions, also addressed by the study of microsatellites and SNPs, can also be employed to ask a quite different question, which is, What *fraction* of all human genetic variation, whether based on protein coding genes, microsatellites, or any other polymorphic DNA sequences, lies within geographically separated populations and what fraction lies between these populations? This is not an assignment problem, but a question of the average *amount* of genetic diversification between and within geographical groups. The two problems can be related to each other by posing the question, Are the genes that are geographically highly differentiated in their allelic frequencies typical of the human genome in general? The answer to that question turns out to be "no." While there are indeed genes whose allelic frequencies differ markedly between geographical regions and can be used for taxonomic purposes, these are not typical of the human genome in general.

The Problem of Ancestry

It has long been known that some loci are highly differentiated between geographical populations. Indeed, one does not need to be a geneticist to solve the taxonomic problem at first sight. Using skin color, facial shape, and hair form, all obviously largely genetically determined (although the genes influencing these characters have only begun to be localized), no one has any difficulty in differentiating between a random person taken from

West Africa, from China, from Norway, and from the tropical rainforest of the Orinoco basin. With only a little more subtlety one can differentiate Amharic-speaking natives of Ethiopia from Zulus, Chinese from Japanese, and villagers of Andhra Pradesh from Afghanis by external morphology. Some classic blood-group polymorphisms are highly differentiated geographically, although most are not (Cavalli-Sforza, Menozzi, & Piazza, 1994). Thus, it is not surprising that DNA sequences such as microsatellite markers can be used to infer the major region of origin of individuals. In the case of the microsatellites, the differences in allelic frequencies between groups are, in fact, small so that data from a very large number of markers had to be subjected to a sophisticated statistical clustering technique to make reliable inferences about geographical origin.

Let us first make clear what has been discovered about ancestry from two recent studies of microsatellites and insertion-deletion polymorphisms from the Human Genome Diversity Cell Line Panel (HGDP-CEPH) collection. In the first study (Rosenberg et al., 2002), 377 microsatellite markers were studied in 1,056 individuals from 52 sites representing native populations from all continents. The second study (Rosenberg et al., 2005) included 783 microsatellite markers and 210 insertion-deletion polymorphisms. There was a slight difference between the samples of individuals used in the two studies, with 1,048 individuals representing 53 populations in the second study. The reasons for this difference have to do with possible duplication of a small number of samples and the inclusion of a small group of Bantu samples into a single group.

For the two studies, the conclusions about continental ancestry are remarkably similar despite the difference in the size of the data sets. The essential finding is that these highly variable markers can be used to form affinity clusters on the basis of similarities between individuals in their genotypes. The statistical technique used (Pritchard, Stevens, & Donnelly, 2000) finds the most probable assignment of individuals to clusters, and this is done blind to knowledge of the actual geographic origin of the individuals.[2] After these clusters are formed, they can then be compared to the actual geographic origins of all individuals. For one of the clustering schemes, with five clusters the result was a close fit of the clusters to continents or subcontinents. Many individuals had ancestry from two or more of the clusters, and some clusters showed a great deal of multiple ancestry, a signature of past migrations or conquests or of the continuity of genetic variation in space (King & Motulsky, 2002). Almost all of the sample of Mozabites from Algeria, for example, belonged both to clusters that corresponded to Eurasia and Africa. And the Altaic-speaking Uygurs of northwestern China showed strong ancestry from East Asia and Eurasia. Europe, West Asia, and South/Central Asia are regions of especially mixed ancestry; the separate linguistic

groups in these areas are difficult to separate genetically, even with 783 markers.

The continental clustering in these large sets of data derives mainly from small differences in allele frequencies at large numbers of markers, not from diagnostic genotypes. This clustering reflects the history of human migrations that began when modern humans left Africa 50,000–100,000 years ago (King & Motulsky, 2002; Excoffier, 2003; Cavalli-Sforza & Feldman, 2003). For those geographical regions such as Europe, West Asia, and South/Central Asia that have a long history of migration and colonization, finer resolution of the clusters is very difficult and will probably require more samples and many more polymorphic markers.

It takes a lot of polymorphic microsatellite markers to produce reliable genetic clusters, 50–150 polymorphisms for reliable assignment of continental ancestry. Even with 993 polymorphisms, it remains difficult to resolve finer subdivisions within continents, especially in Europe and Asia. We can conclude, however, that in most cases, self-reported ancestry coincides with the broad continental clustering seen from the genetic markers that were used.

It has been claimed (Serre & Pääbo, 2004) that the geographical clustering seen in the studies referred to above (Rosenberg et al., 2002; Rosenberg et al., 2005) is an artifact of the geographic pattern of samples in the particular data set studied, the HGDP-CEPH (CEPH, n.d.). We had remarked on the existence of "continuous gradients across regions or admixture of neighboring groups" (Rosenberg et al., 2002, p. 2382). In fact, geographic clustering and spatial gradients are both features of these large data sets (Rosenberg et al., 2005). For population pairs from the same cluster, as geographic distance increases, genetic distance increases linearly, consistent with a clinal[3] structure. But for pairs of populations from different clusters, genetic distance is generally larger than between pairs of populations from the same clusters that have the same genetic distance. This suggests that the clusters are formed by the small discontinuous jumps in genetic distance caused by major geographic barriers: oceans, mountain ranges, or deserts. Indeed, the history of migration is important (Ramachandran et al., 2005), but migration is not geographically uniform.

While the great majority of the DNA markers used to define these continental clusters show only small allelic frequency differences between populations, some genes do have greater frequency differences among populations or continents (Bamshad & Wooding, 2003). Duffy and Rh, two of the genes in the original survey of within- and between-population diversity (Lewontin, 1972), show more variation among populations than most blood group or protein genes, microsatellites, or SNPs. The presence of hemoglobin S (causing sickle cell disease) or G6PD (causing favism) in an individual

markedly increases the likelihood that that person has ancestors from a geographic region where malaria was present, while an individual carrying the Tay-Sachs allele is most likely to have Ashkenazi Jewish or French Canadian ancestry. These are cases where populational rather than continental ancestry is the relevant dimension for the allelic differences.

Finally, it must be borne in mind that the taxonomic problem cannot be inverted. That is, while clustering methods are capable of assigning an individual to a geographic population with a high degree of certainty, given that individual's genotype, it is not possible to predict accurately the genotype of an individual given his or her geographical origin. Thus, knowing an individual's ancestry only slightly improves the ability to predict his or her genotype. The more polymorphic the markers, the more difficult this is. This is illustrated in figure 5-1. There are gene alleles that appear only in one group, as for example the Fy^b which is present only in individuals with some European ancestry, but there does not exist any gene for which one major geographical cluster includes 100% of one genotype while another major geographical cluster has 100% of another genotype. Even when the explicit purpose of studies has been to identify markers that show strong differentiation between groups, none that show a complete difference between major groups has been found. In the microsatellite study mentioned earlier (Rosenberg et al., 2002), the most geographically informative loci in the data set have some striking differences, as shown in figure 5-1, but nowhere near 100%. In figure 5-1, the size of the pie slice with a given degree of grayness represents its frequency in the region specified by the column. In some regions, some alleles are rare and do not occupy enough area to be seen. The top marker (D12S2070) shows very different allele frequencies in the different regions, the middle one (D10S1425) shows moderate frequency differences, and the bottom marker (D6S474) shows very small differences. All three loci have eight alleles, which are shown in increasing order of allele size (i.e., number of species) in a counterclockwise manner, starting from the top of each circle (for each locus, the smallest allele is shown in white, the largest is shown in black).

The Problem of Allocation of Variation

When we turn from the problem of finding genes that will discriminate ancestry to the problem of the relative amount of human genetic diversity that lies within and between populations, there is no controversy. The first survey, in 1972, of genetic diversity over a very large sample of local human populations from major geographical regions used all the available data for blood groups and enzyme proteins for every local human population that had been studied up to that time (Lewontin, 1972). The result was that 85%

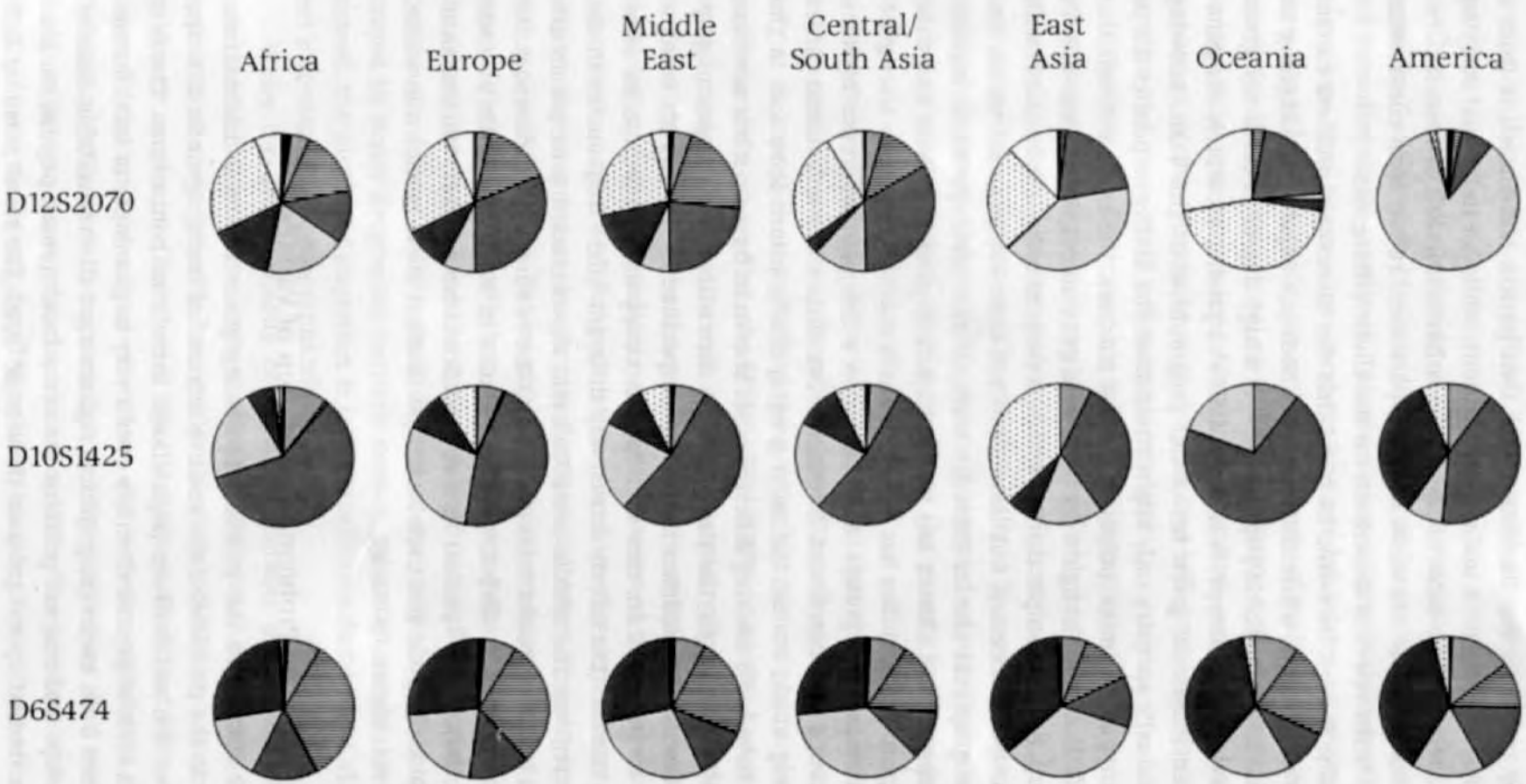

FIGURE 5-1. Distributions of allele frequencies across seven regions for three microsatellite markers.

From Rosenberg et al. (2002).

of all human genetic diversity, measured by the Shannon-Weaver information measure (or its close equivalent, heterozygosity) is present within local national groups, that is, averaging within Swedes, within Kikkuyu, within Japanese, etc. An additional 8% is present between local groups within what were designated classically as races, between Swedes, Italians, and Greeks, or between Kikkuyu, Zulu, and Hutu. The remaining 7% lies between the classical major races, between sub-Saharan Africans, East Asians, Australian Aborigines, Europeans, etc. Several similar studies were subsequently carried out on smaller geographical samples with similar results for the within-population variation, but with roughly 5% of the variation between local populations and 10% among the major "races." When these studies were repeated using a limited amount of DNA sequence variation (Barbujani, Magagni, Minch, & Cavalli-Sforza, 1997), again 85% of the variation was found within local populations, with about 5% between local populations and 10% among major classical races. The studies of the DNA markers discussed in the previous section (Rosenberg et al., 2002; Rosenberg et al., 2005) in the context of the taxonomic problem also partitioned total variation using a different measure of diversity. With this measure, 86% to 95% of the diversity was assigned within local populations, between 2% and 6% among populations within major geographical regions and between 3% and 10% among major regions (classical races).

The higher than usual estimate of between 93% and 95% for the within-population component of genetic variation arises from several sources. The previous studies used samples from isolated and geographically well-separated populations. On restriction of the microsatellite analysis to populations chosen to mimic the pattern in these previous studies, the within-population component in the microsatellite study was reduced to 89.8%. A further contribution to the difference is from the greater heterozygosity of microsatellites, of which only 30 were included in the 109 genes studied by Barbujani et al. Differential selection on protein variants across geographical regions might also augment the between-population component as compared to the microsatellite study. The different studies differ on the exact partition of variance. Nevertheless, the strong overall conclusion is that although it is possible to use genetic divergences to assign individuals to regions of origin with high confidence, there is very little average genetic difference among geographical regions as compared to the variation observed within any local population.

The Problem of Race

Race as a biological concept has had a variety of disparate meanings, even within the last 50 years. In the classical taxonomic literature a race was any

distinguishable type within a species, as, for example, dark-bellied and light-bellied races of small mammals. While such races sometimes corresponded to geographically separated populations, phenotypic differences on which racial classifications were based were often a consequence of single gene differences, so that two siblings could be of different "race." In reaction against this typological notion of race, Dobzhansky introduced the notion of "geographical races" which were defined as "populations of species that differ in the frequencies of one or more genetic variants, gene alleles or chromosomal structures" (1937, p. 138). The problem with this definition is that every geographical population of every species in the world is a "geographical race" because no two populations have identical allelic frequencies for polymorphic genes, so geographical race becomes synonymous with population.

The typological and the geographic notions of race are combined in the classical division of human races because it is observed that the native inhabitants of different major regions of the world are characterized by clear phenotypic differences of color, facial features, and hair form. Variation in these phenotypes is also observed among individuals within races to the extent that even categorization according to such traits can be difficult (Brown & Armelagos, 2001). An underlying assumption of human race classification, a classification based on a small number of obvious phenotypic differences, was that these differences were characteristic of the genome in general. Just as there were large differences in genes for color, so there would be large differences in genes influencing cognitive and most physiological traits. Indeed, in the absence of any evidence to the contrary, this is not an absurd assumption, but it turns out to be wrong. The repeated and consistent results on the apportionment of genetic diversity reviewed in the previous section show that the genes underlying the phenotypic differences used to assign race categories are atypical of the genome in general and are not a reliable index to the amount of genetic differentiation between groups. Thus, racial assignment loses any general biological interest. For the human species, race assignment of individuals does not carry with it any general implication about genetic differentiation.

With the advent of large movements of populations between continents, especially with European colonial expansion and the commercial slave trade, new populations have arisen which are mixtures of the major continental groups, especially in the Western Hemisphere and Oceania. Large numbers of people, then, have ancestry from more than one major geographical region so that the association of phenotype and geography breaks down and race again becomes a typology with even less power to distinguish the genomes of those involved. What is the race of a dark-skinned person, half of whose ancestry is of sub-Saharan African origin and half of Northern European

origin? Social practice in the United States makes an asymmetrical nominal assignment of race such that any detectable African ancestry makes a person "black" or "African American," but this obfuscates the biological reality.

What do the clusters constructed, for example, from the data in the microsatellite studies have to do with our common understanding of race? We must remember that the clusters are defined by markers that have no influence on obvious phenotypes. Nevertheless, there are some phenotypes that correlate well with continental origin, usually those that we can see. But even here we can be misled—dark skin is a feature of sub-Saharan Africans but also of southern Indians and Australian aborigines. Thus, if we were to use skin color alone, continental clustering, i.e., common racial classification, fails.

The Use of Race and Ancestry in Medicine

It is often claimed that racial categorization is of considerable importance in medicine because there are a number of loci of medical relevance that are highly differentiated between geographical populations.

A focus of recent studies on the medical relevance of race and/or ethnicity concerns variation in drug metabolizing enzymes, DMEs (Yancy et al., 2001; Exner, Dries, Domanski, & Cohen, 2001; Xie, Kim, Wood, & Stein, 2001; Schwartz, 2001; Wilson et al., 2001; Risch, Burchard, Ziv, & Tang, 2002). Some of these enzymes appear to differ in allele frequencies among Americans of different ethno-cultural backgrounds (Yancy et al., 2001; Exner et al., 2001; Xie et al., 2001). Wilson and his group compared differences in DME frequencies among genetically estimated clusters obtained using 39 microsatellites assayed on individuals from eight populations with corresponding differences among the same individuals classified by ethnicity. They claim that the ethnic labels are "insufficient and inaccurate" surrogates for the genetic clusters and are less valuable than the latter in resolving group-specific profiles of DMEs. This finding is contested by Risch and colleagues (2002), who go further to claim that the use of genetic clusters instead of a racial classification may cause the effects of socioeconomic, environmental, and lifestyle variation on a disease to be underestimated.

For DMEs, as we know for blood groups and other enzymes, it is reasonable to predict that variation within will far outweigh that between continental groups. For these genes, too, only variation at a much finer level than continents or races may provide information about ancestry that is phenotypically relevant.

The situation is even more complicated when we examine diseases that appear to aggregate in the classically defined races. Sickle cell disease is one that is often thought to be an African trait. But it exists in a number of

Mediterranean and Indian populations as well. Sickle cell is not a marker of skin color or race, but more properly a marker of ancestry in a geographic location where malaria is or was prevalent. And, of course, not all Africans or Sardinians carry the gene responsible for sickle cell disease. Thus, classical race is not diagnostic of the disease, and the disease is not diagnostic of race. Rosenberg et al.'s clusters (Rosenberg et al., 2002; Rosenberg et al., 2005) don't tell us very much about traits that are determined by genes that have been under selection. It is nevertheless the case that a knowledge of ancestry can play an important role in medical diagnosis and drug therapy. Thus, we might make a classification based only on the sickle cell phenotype, or the Tay-Sachs phenotype, or lactose intolerance. From a medical point of view, a breakdown of humans into the hundreds if not thousands of such groups might say more about the biology of disease than mere continental ancestry. Knowledge of ancestry with respect to these subpopulations may then be informative about the risks of disease.

Both the regional heterogeneity within major geographical regions and the widespread mixture of formerly relatively isolated populations result in a confusion between race and ancestry that is critical and must be accounted for in medical practice. The assignment of racial classification to an individual hides the biological information that is needed for intelligent therapeutic and diagnostic decisions (see also Tate & Goldstein, this volume). A person classified as "black" or "Hispanic" by social convention may have any mixture of European, African, Native American, and, more rarely, Asian ancestry. Moreover, there is genetic heterogeneity among regions within these major geographical groups. If we have a serious interest in making diagnostic and therapeutic decisions based on genotype, then it is not typological race assignment that is relevant but the various contributions to a person's ancestry that are informative. The kind of questions to be asked are these: Do you have any African ancestors? If so, do you know from what part of Africa they came? Do you have any European ancestry? If so, from what part of Europe did they come? Were there any Ashkenazi Jews among your ancestors? And so on. The detailed information about local geographical origins will often be unavailable, but categorical racial assignments are not a substitute for some kind of more informative ancestry history.

We agree entirely with Risch et al. (2002) that conventional socially defined race which, for example, classifies all persons with visually detectable African ancestry as "black" or "African American" is of use in a medical context to the extent that it provides information about social circumstances and lifestyle conditions of patients, particularly discrimination. But these socially defined categories should not be confounded with genetically defined races. The actual distribution of human genetic variation, including the distribution of genotypes that are directly relevant to the diagnosis

and treatment of disease, is such that race is not a useful biological concept when applied to humans. It is nevertheless true that data about the various lines of ancestry of an individual can provide information on the likelihood that the person carries certain gene alleles. Lines of ancestry, rather than genetically arbitrary racial categories, can provide much accurate, biologically interesting, and potentially medically useful information. For diagnosis and treatment, however, individual genotypes will, in the long run, provide the most useful information.

NOTES

1. A microsatellite is any of numerous short segments of DNA that are distributed throughout the genome, that consist of repeated sequences of usually two to five nucleotides, and that tend to vary from one individual to another.
2. See Bolnick (this volume) for a discussion of Pritchard's computer program, *structure*.
3. Clinal genetic variation refers to a gradient of change in a group of related organisms, usually along a line of environmental or geographic transition.

REFERENCES

Bamshad, M., & Wooding, S. P. (2003). Signatures of natural selection in the human genome. *Nature Reviews Genetics, 4*, 99–111.

Bamshad, M. J., Wooding, S. P., Watkins, W. S., Ostler, C. T., Batzer, M. A., & Jorde, L. B. (2003). Human population genetic structure and inference of group membership. *American Journal of Human Genetics, 72*, 578–589.

Barbujani, G., Magagni, A., Minch, E., & Cavalli-Sforza, L. L. (1997). An apportionment of human DNA diversity. *Proceedings of the National Academy of the Sciences USA, 94*, 4516–4519.

Bolnick, D. A. (2008 [this volume]). Individual ancestry inference and the reification of race as a biological phenomenon. In B. A. Koenig, S. S.-J. Lee, & S. S. Richardson (Eds.), *Revisiting race in a genomic age* (pp. 70–85). New Brunswick, NJ: Rutgers University Press.

Brown, R. A., & Armelagos, G. J. (2001). Apportionment of racial diversity: A review. *Evolutionary Anthropology, 10*, 34–40.

Burchard, E. G., Ziv, E., Coyle, N., Gomez, S. L., Tang, H., Karter, A. J., et al. (2003). The importance of race and ethnic background in biomedical research and clinical practice. *New England Journal of Medicine, 348*, 1170–1175.

Calafell, F. (2003). Classifying humans. *Nature Genetics, 33*, 435–436.

Cavalli-Sforza, L. L., & Feldman, M. W. (2003). The application of molecular genetic approaches to the study of human evolution. *Nature Genetics, 33* (Suppl.), 266–275.

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes*. Princeton, NJ: Princeton University Press.

Centre d'Etude du Polymorphisme Humain (CEPH) (n.d.). HGDP-CEPH Human Genome Diversity Cell Line Panel. Retrieved August 7, 2006, from http://www.cephb.fr/HGDP-CEPH-Panel/.

Conrad, D. F., Jakobsson, M., Coop, G., Wen, X., Wall, J. D., Rosenberg, N. A., et al. (2006). A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nature Genetics, 38*, 1251–1260.

Cooper, R. S., Kaufman, J. S., & Ward, R. (2003). Race and genomics. *New England Journal of Medicine, 348*, 1166–1170.

Dobzhansky, T. (1937). *Genetics and the origin of species*. New York: Columbia University Press.

Excoffier, L. (2003). Human diversity: Our genes tell where we live. *Current Biology, 13*, R134–R136.

Exner, D. V., Dries, D. L., Domanski, M. J., & Cohen, J. N. (2001). Lesser response to angiotensin-converting-enzyme inhibitor therapy in black as compared with white patients with left ventricular dysfunction. *New England Journal of Medicine, 344*, 1351–1357.

Genes, drugs, and race. (2001). *Nature Genetics, 29*, 239–240.

King, M., & Motulsky, A. G. (2002). Mapping human history. *Science, 298*, 2342–2343.

Lewontin, R. C. (1972). The apportionment of human diversity. *Evolutionary Biology, 6*, 381–398.

Phimister, E. G. (2003). Medicine and the racial divide. *New England Journal of Medicine, 348*, 1081–1082.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*, 945–959.

Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W., & Cavalli-Sforza, L. L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of the Sciences USA, 102*, 15942–15947.

Risch, N., Burchard, E., Ziv, E., & Tang, H. (2002). Categorization of humans in biomedical research: Genes, race and disease. *Genome Biology, 3*, 2007.1–2007.12.

Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., & Feldman, M. W. (2005). Clines, clusters, and the effect of study design on the inference of human population structure. *PloS Genetics, 1*, 660–671.

Rosenberg, N. A., Pritchard, J. K., Cann, H., Weber, J., Kidd, K. K., Zhivotovsky, L. A., et al. (2002). Genetic structure of human populations. *Science, 298*, 2381–2385.

Sankar, P., & Cho, M. K. (2002). Toward a new vocabulary of human genetic variation. *Science, 298*, 1337–1338

Satel, S. (2002, May 5). I am a racially profiling doctor. *The New York Times*, sec. 6, p. 56.

Schwartz, R. S. (2001). Racial profiling in medical research. *New England Journal of Medicine, 344*, 1392–1393.

Serre, D., & Pääbo, S. (2004). Evidence for gradients of human genetic diversity within and among continents. *Genome Research, 14*, 1679–1685.

Stephens, J. C., Schneider, J. A., Tanguay, D. A., Choi, J., Acharya, T., Stanley, S. E., et al. (2001). Haplotype variation and linkage disequilibrium in 313 human genes. *Science, 293*, 489–493.

Tate, S. K., & Goldstein, D. B. (2008 [this volume]). Will tomorrow's medicines work for everyone? In B. A. Koenig, S. S.-J. Lee, & S. S. Richardson (Eds.), *Revisiting race in a genomic age* (pp. 102–128). New Brunswick, NJ: Rutgers University Press.

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology, 4*, e72.

Wade, N. (2002a, July 30). Race is seen as real guide to track roots of disease. *New York Times*, Science sec., p. 1.

Wade, N. (2002b, December 20). Gene study identifies f main human populations, linking them to geography. *New York Times*, National sec., p. 29.

Wade, N. (2003, March 20). 2 scholarly articles diverge on role of race in medicine. *New York Times*, Late ed.–final, sec. A, p. 30.

Wilson, J. F., Weale, M. E., Smith, A. C., Gratrix, F., Fletcher, B., Thomas, M. G., et al. (2001). Population genetic structure of variable drug response. *Nature Genetics, 29*, 265–269.

Wood, A.J.J. (2001). Racial differences in the response to drugs—pointers to genetic differences. *New England Journal of Medicine, 344*, 1393–1395.

Xie, H. G., Kim, R. B., Wood, A. J., & Stein, C. M. (2001). Molecular basis of ethnic differences in drug disposition and response. *Annual Review of Pharmacology and Toxicology, 41*, 815–850.

Yancy, C. W., Fowler, M. B., Colucci, W. S., Gilbert, E. M., Bristow, M. R., Coln, J. N., et al. (2001). Race and the response to adrenergic blockade with carvedilol in patients with chronic heart failure. *New England Journal of Medicine, 344*, 1358–1365.