nature
genetics

Genetic variation, classification and 'race'

Lynn B Jorde & Stephen P Wooding

New genetic data has enabled scientists to re-examine the relationship between human genetic variation and 'race'. We review the results of genetic analyses that show that human genetic variation is geographically structured, in accord with historical patterns of gene flow and genetic drift. Analysis of many loci now yields reasonably accurate estimates of genetic similarity among individuals, rather than populations. Clustering of individuals is correlated with geographic origin or ancestry. These clusters are also correlated with some traditional concepts of race, but the correlations are imperfect because genetic variation tends to be distributed in a continuous, overlapping fashion among populations. Therefore, ancestry, or even race, may in some cases prove useful in the biomedical setting, but direct assessment of disease-related genetic variation will ultimately yield more accurate and beneficial information.

Few concepts have as tarnished and contentious a history as 'race'[1]. Among both the scientific and lay communities, the notion that humans can be grouped into different races has been enshrined by some and dismissed by others. Even the definition of race varies considerably, depending on context and criteria[2,3]. Nevertheless, race continues to be used in a variety of applications. Forensic databases in the US are typically organized according to traditional racial and ethnic categories (*e.g.*, African-American, European-American, Hispanic). Investigators funded by the US National Institutes of Health are required to show that minority populations are adequately represented in biomedical studies. Responses to medical therapies, such as drugs, are often compared among populations that are divided according to traditional racial divisions. Among the general public, the validity of racial categories is often taken for granted.

Not surprisingly, biomedical scientists are divided in their opinions about race. Some characterize it as "biologically meaningless"[4] or "not based on scientific evidence"[5], whereas others advocate the use of race in making decisions about medical treatment or the design of research studies[6–8]. Amid this controversy, modern human genetics has generated a staggering array of new data. For the first time, it is possible to study human genetic variation using not just a few dozen polymorphisms, but hundreds or even thousands. In addition to neutral polymorphisms that inform us about population history, increasing numbers of variants that contribute to disease are being discovered.

Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA. Correspondence should be addressed to L.B.J. (lbj@genetics.utah.edu).

In this review, we provide a general overview of patterns of human variation, first at the population level and then at the individual level. We also discuss whether current genetic data support traditional concepts of race, and the implications of these data for medical research and practice.

Variation at the population level

The average proportion of nucleotide differences between a randomly chosen pair of humans (*i.e.*, average nucleotide diversity, or π) is consistently estimated to lie between 1 in 1,000 and 1 in 1,500 (refs. 9,10). This proportion is low compared with those of many other species, from fruit flies to chimpanzees[11,12], reflecting the recent origin of our species from a small founding population[13]. The π value for *Homo sapiens* can be put into perspective by considering that humans differ from chimpanzees at only 1 in 100 nucleotides, on average[14,15]. Because there are approximately three billion nucleotide base pairs in the haploid human genome, each pair of humans differs, on average, by two to three million base pairs.

Of the 0.1% of DNA that varies among individuals, what proportion varies among main populations? Consider an apportionment of Old World populations into three continents (Africa, Asia and Europe), a grouping that corresponds to a common view of three of the 'major races'[16,17]. Approximately 85–90% of genetic variation is found within these continental groups, and only an additional 10–15% of variation is found between them[18–20] (**Table 1**). In other words, ~90% of total genetic variation would be found in a collection of individuals from a single continent, and only ~10% more variation would be found if the collection consisted of Europeans, Asians and Africans. The proportion of total genetic variation ascribed to differences between continental populations, called F_{ST}, is consistent, regardless of the type of autosomal loci examined (**Table 1**). F_{ST} varies, however, depending on how the human population is divided. If four Old World populations (European, African, East Asian and Indian subcontinent) are examined instead of three, F_{ST} (estimated for 100 *Alu* element insertion polymorphisms) decreases from 14% to 10% (ref. 21). These estimates of F_{ST} and π tell us that humans vary only slightly at the DNA level and that only a small proportion of this variation separates continental populations.

Allele frequencies in populations can be compared to assess the extent to which populations differ from one another. The overall pattern of these genetic distances, summarized in a network diagram (**Fig. 1**), shows several important trends. First, populations tend to cluster according to their geographic distance from one another. This is to be expected, as geographically distant populations were less likely to exchange migrants throughout human evolutionary history. Second, the African populations are more diverse, a pattern consistent with

Table 1 Distribution of genetic variation in Old World continental populations

	60 STR polymorphisms	30 restriction site polymorphisms	100 *Alu* insertion polymorphisms
Between individuals, within continents	90%	87%	86%
Between continents	10%	13%	14%

Data are from refs. 18,21. Old World continents are Africa, Asia and Europe. Similar values are obtained when estimates are based on groups from additional regions, such as Oceania and the New World[19,71].

many studies that have compared π in various human populations[22–24]. Third, the largest genetic distance is seen between African and non-African populations. Finally, because this network is based on *Alu* insertion polymorphisms, it is possible to designate unambiguously the allelic state of the ancestral human population as one in which none of the insertions would be present. This allows us to 'root' the tree and shows that the root falls closest to African populations. All of these findings, which are in accord with many other studies based on different types of genetic variation assessed in different samples of humans[25–28], support an evolutionary scenario in which anatomically modern humans evolved first in Africa, accumulating genetic diversity. A small subset of the African population then left the continent, probably expe-

rienced a population bottleneck and founded anatomically modern human populations in the rest of the world[29]. Of special importance to discussions of race, our species has a recent, common origin.

Variation at the individual level

Comparisons of populations are sometimes criticized because the allocation of individuals into groups imposes a pre-existing structure and may influence the outcome of a genetic study. Furthermore, populations are defined in many (often arbitrary) ways. Some of these objections can be overcome by comparing individuals rather than populations.

A common approach in studying individual genetic variation is to compute the genetic similarity between all possible pairs of individuals in a sample and then to search for clusters of individuals who are most similar to one another. At least two such studies have concluded that the observed clustering patterns do not correspond well to the subjects' geographic origins or 'ethnic labels'[30,31]. These studies, however, were based on only several dozen or fewer loci, and a small sample of unselected loci does not typically provide sufficient power to detect population structure when individuals are analyzed[6,32]. In contrast, studies based on more loci[32–34] show that individuals tend to cluster according to their ancestry or geographic origin. **Figure 2** shows a tree in which genetic similarity, based on 190 loci, is portrayed among individual members of most of the populations illustrated in **Figure 1**. The longest branches in this tree separate individuals within the same continental populations, as expected from the F_{ST} results discussed above (*i.e.*, most variation occurs within populations). The longest internal branch separates African from non-African individuals, again in agreement with previous results at the population level. The next cluster consists entirely of Europeans, and a final cluster contains all of the East Asian subjects and one European. The robustness and validity of these findings are supported by other studies, which, despite using different loci and different population samples, obtained similar patterns[34–37].
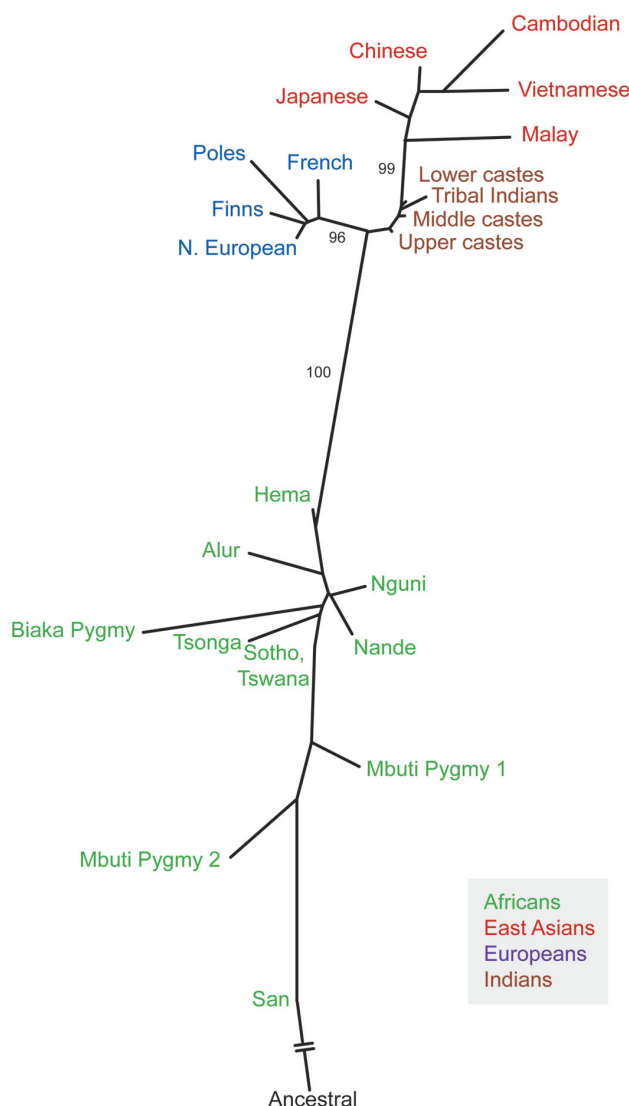

Figure 1 A neighbor-joining network of population similarities, based on the frequencies of 100 *Alu* insertion polymorphisms. The network is rooted using a hypothetical ancestral group that lacks the *Alu* insertions at each locus. Bootstrap values are shown (as percentages) for main internal branches. (Because of the relatively small sample sizes of some individual populations, bootstrap values for terminal branches within main groups are usually smaller than those of the main branches, indicating less statistical support for terminal branches.) The population groups and their sample sizes are as follows: Africans (152): Alur, 12; Biaka Pygmy, 5; Hema, 18; Coriell Mbuti Pygmy, 5; a second sample of Mbuti Pygmy from the Democratic Republic of the Congo, 33; Nande, 17; Nguni, 14; Sotho/Tswana, 22; Kung (San), 15; Tsonga, 14. East Asians (61): Cambodian, 12; Chinese, 17; Japanese, 17; Malay, 6; Vietnamese, 9. Europeans (118): northern Europeans, 68; French, 20; Poles, 10; Finns, 20. South Indians (365): upper caste Brahmin, Kshatriya and Vysya, 81; middle caste Kapu and Yadava, 111; lower caste Relli, Mala and Madiga, 74; tribal Irula, Khonda Dora, Maria Gond and Santal, 99. Further details on samples and methods are given in ref. 21.

A statistically more sophisticated approach for cluster definition is afforded by the structure program[38], in which individuals are first randomly assigned into one of k groups, without regard to population affiliation. Individuals are then moved between groups to minimize Hardy-Weinberg disequilibrium and interlocus linkage disequilibrium within each group (minimal within-groups disequilibrium ensures genetic homogeneity within each group). Posterior probabilities of membership in each of the k groups are then estimated for each individual. **Figure 3a** shows the results of applying the structure program to individuals from the populations shown in **Figure 2**. When 100 *Alu* insertion polymorphisms and 60 short tandem repeat (STR) polymorphisms were used, all Europeans, East Asians and Africans were correctly placed according to their respective continents of origin[32].

These results may seem paradoxical in light of the small proportion of genetic variation that separates continental populations (F_{ST}), but F_{ST} should not be confused with the accuracy of classification. Incorporation of additional loci in a population study has minimal effects on F_{ST} but steadily increases the accuracy of classification[39]. This is because F_{ST} treats each locus in isolation from the others; it takes no account of correlations among loci. The more loci there are in the data set, the more of these correlations there are, and the more information is ignored by F_{ST}.

Considering the results shown in **Figures 2** and **3a**, it might be tempting to conclude that genetic data verify traditional concepts about races. But the individuals used in these analyses originated in three geographically discontinuous regions: Europe, sub-Saharan Africa and East Asia. When a sample of South Indians, who occupy an intermediate geographic position (see also **Fig. 1**) is added to the analysis (**Fig. 3b**), considerable overlap is seen among these individuals and both the East Asian and European samples, probably as a result of numerous migrations from various parts of Eurasia into India during the past 10,000 years[40]. Thus, the South Indian individuals do not fall neatly into one of the categories usually conceived as a 'race'. In addition, examination of the posterior probabilities estimated by structure shows that most individuals in **Figure 3** are not classified with 100% probability into one of the main clusters[32] (similar results were obtained in ref. 33). In other words, each individual within a cluster shares most, but not all, of his or her ancestry with other members of the cluster (*e.g.*, a member of the European cluster might have a posterior probability of 90% for assignment to that cluster, with 5% probability of assignment into each of the other two clusters in **Fig. 3a**). Ancestry, then, is a more subtle and complex description of an individual's genetic makeup than is race[41]. This is in part a consequence of the continual mixing and migration of human populations throughout history. Because of this complex and interwoven history, many loci must be examined to derive even an approximate portrayal of individual ancestry.

The picture that begins to emerge from this and other analyses of human genetic variation is that variation tends to be geographically structured, such that most individuals from the same geographic region will be more similar to one another than to individuals from a distant region. Because of a history of extensive migration and gene flow, however, human genetic variation tends to be distributed in a continuous fashion and seldom has marked geographic discontinuities[19,42]. Thus, populations are never 'pure' in a genetic sense, and definite boundaries between individuals or populations (*e.g.*, 'races') will be necessarily somewhat inaccurate and arbitrary.

Though now supported by many genetic data, this concept is hardly new. Blumenbach[43], writing in the 1700s, acknowledged extensive morphological overlap among populations or races. Charles Darwin, some 100 years later, wrote, "It may be doubted whether any character can be named which is distinctive of a race and is constant."[44] Scientists recognized shared genetic variation among populations more than 50 years ago, although their conclusions were based on relatively small numbers of informative loci[45]. This pattern of shared variation has important implications for our understanding of population differences and similarities, and it also bears on critical biomedical issues.
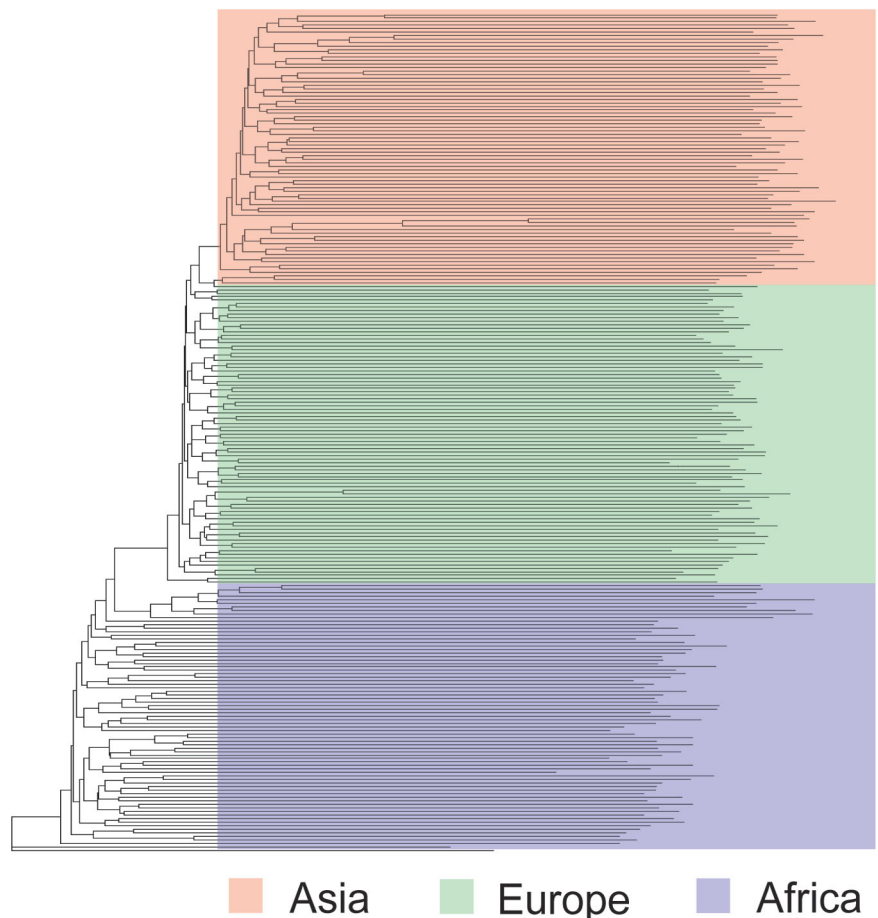

Asia   Europe   Africa

Figure 2 A neighbor-joining tree of individual similarities, based on 60 STR polymorphisms[72], 100 *Alu* insertion polymorphisms[21] and 30 restriction site polymorphisms[73]. The percentage of shared alleles was calculated for all possible pairs of individuals, and a neighbor-joining tree was formulated using the PHYLIP software package[74]. African individuals are shown in blue, European individuals in green and Asian individuals in orange.
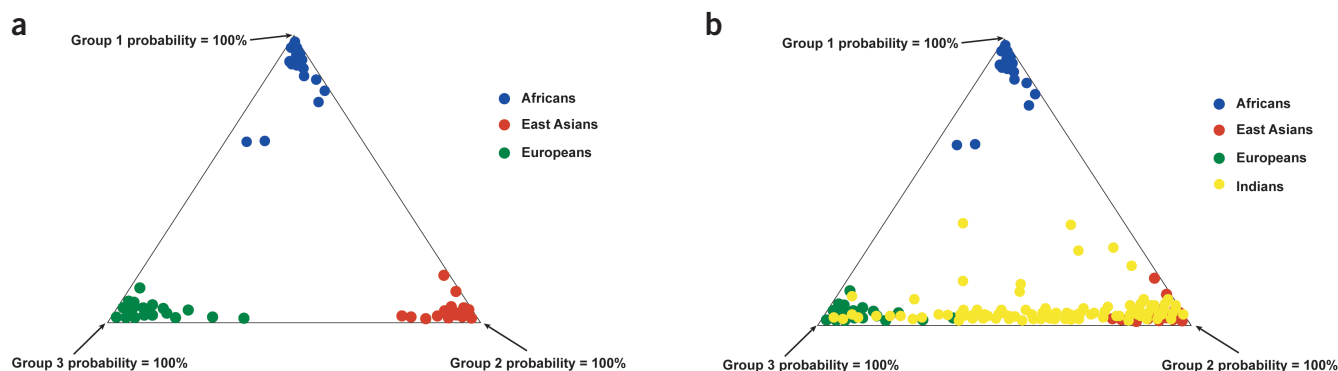
Figure 3 Structure analysis. (**a**) Results of applying the structure program to 100 *Alu* insertion polymorphisms typed in 107 sub-Saharan Africans, 67 East Asians and 81 Europeans. Individuals are shown as dots in the diagram. Three clusters appear in this diagram; a cluster membership posterior probability of 100% would place an individual at an extreme corner of the diagram. (**b**) A second application of the structure program, using the individuals shown in **a** as well as 263 members of caste populations from South India. Adapted from ref. 32.

Genetic variation, race and medicine

Race and ethnicity have long been incorporated into medical decision-making processes. For example, physicians are typically aware that sickle-cell disease is much more common in African and Mediterranean populations than in northern European populations, whereas the reverse is true for cystic fibrosis and hemochromatosis. Although such distinctions are often clearest for single-gene diseases, perceived population differences influence the diagnosis and treatment of common diseases as well. There is evidence, for example, for population differences in response rates to drugs used in treating hypertension[46] and depression[47,48].

Broad population categories can be discerned genetically when enough polymorphisms are analyzed, as seen in **Figure 3**, so these categories are not devoid of biological meaning. When several thousand or more polymorphisms are examined, individual populations, such as Japanese and Chinese, can be delineated[34], and members of 'admixed' American populations, such as Hispanics, African-Americans and European-Americans, can be accurately identified[34,49]. Similar results are obtained whether coding or noncoding polymorphisms are used[49].

At face value, such results could be interpreted as supporting the use of race in evaluating medical treatment options. But race and ancestry are not equivalent. Many polymorphisms are required to estimate an individual's ancestry, whereas the number of genes involved in mediating a specific drug response may be relatively small[50]. If disease-associated alleles are common (and thus of clinical significance), they are likely to be relatively ancient and therefore shared among multiple populations[51,52]. Consequently, an individual's population affiliation would often be a faulty indicator of the presence or absence of an allele related to diagnosis or drug response.

To illustrate these points, consider the gene angiotensinogen (*AGT*), which encodes a key component of the renin-angiotensin blood-pressure regulation pathway. An *AGT* variant, 235T, is associated (in many populations) with a 10–20% increase in the risk of develop-
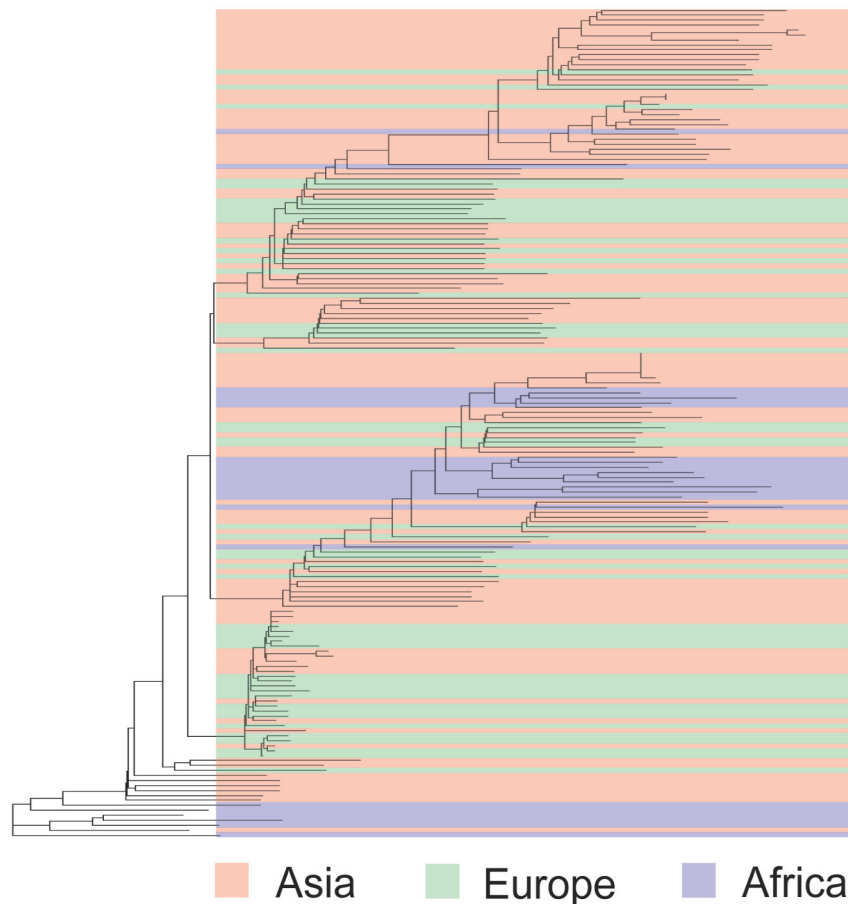

Asia   Europe   Africa

Figure 4 A neighbor-joining tree formulated using the same methods as in **Figure 2**, based on polymorphisms in the 14.4-kb gene *AGT*. A total of 246 sequence variants, including 100 singletons, were observed. The 368 European, Asian and African individuals are described further in ref. 54.

ing hypertension[53]. This variant has a frequency as high as 90% in some African populations and as low as 30% in European populations[54]. There is substantial population variation in the frequency of this allele, but it is present with an appreciable frequency in all populations examined to date. If a tree is made using allelic variation in this 14.4-kb gene (**Fig. 4**), the results differ markedly from those shown in **Figure 2**. In many cases, individuals from different continents are more similar to one another, with regard to this gene, than are individuals from the same continent.

Another example is given by the gene *CYP2D6*, which encodes a member of the cytochrome P450 family involved in the metabolism of many important drugs[55]. Null alleles of *CYP2D6* render the gene product inactive. Individuals who are homozygous with respect to *CYP2D6* null alleles cannot, for example, effectively convert codeine to its active form, morphine, and experience little or no analgesic effect. The median frequencies of null *CYP2D6* alleles vary from 6% in Asian populations to 7% in African populations and 26% in European populations[56]. Thus, there is substantial population variation, but these variants are present in all populations. Population affiliation alone would be, at best, a crude and potentially inaccurate indicator of response to codeine and other CYP2D6-metabolized drugs.

Much discussion has been devoted to recent findings that hypertensive African-Americans show less response, on average, than hypertensive European-Americans to angiotensin-converting enzyme (ACE) inhibitors[46,57]. A recent meta-analysis showed that the average difference in systolic blood pressure reduction in African-Americans versus European-Americans was only 4.6 mm Hg, and the standard deviations of the change in blood pressure in European-Americans and African-Americans were 12 and 14 mm Hg, respectively[58]. Clearly, many African-Americans would respond better to ACE inhibitors than would many European-Americans. To conclude, on the basis of population averages, that ACE inhibitors are ineffective in African-Americans could deny many people a powerful and appropriate drug treatment.

Patterns such as these are seen in many genes that are thought to underlie susceptibility to common diseases. Allelic variation tends to be shared widely among populations, so race will often be an inaccurate predictor of response to drugs or other medical treatments. It would be far preferable to test directly the responsible alleles in affected individuals[50,59]. Such an assessment of allelic variation is a hope for the future, because most of the genes that contribute to common diseases remain to be identified. In addition, nongenetic factors nearly always have an important (and sometimes predominant[60]) role in disease susceptibility. Genetic assessment alone will never be a panacea. Nevertheless, progress is being made in identifying alleles that underlie complex disease susceptibility[61–65], and rapid developments in new technologies such as microarrays promise to provide efficient and far-ranging genetic assays[66].

Conclusions

Data from many sources have shown that humans are genetically homogeneous and that genetic variation tends to be shared widely among populations. Genetic variation is geographically structured, as expected from the partial isolation of human populations during much of their history. Because traditional concepts of race are in turn correlated with geography, it is inaccurate to state that race is "biologically meaningless." On the other hand, because they have been only partially isolated, human populations are seldom demarcated by precise genetic boundaries. Substantial overlap can therefore occur between populations, invalidating the concept that populations (or races) are discrete types.

When large numbers of loci are evaluated, it is often possible to infer individual ancestry, at least approximately. If done accurately and with appropriate reservations, ancestral inference may be useful in genealogical studies, in the forensic arena and in the design of case-control studies. This should not be confused, however, with the use of ethnicity or race (genetically measured or self-identified) to make decisions about drug treatment or other medical therapies. Responses to these therapies will often involve nongenetic factors and multiple alleles, and different populations will often share these alleles. When it finally becomes feasible and available, individual genetic assessment of relevant genes will probably prove more useful than race in medical decision making.

In the meantime, ethnicity or race may in some cases provide useful information in biomedical contexts, just as other categories, such as gender or age, do. But the potential usefulness of race must be balanced against potential hazards. Ignorance of the shared nature of population variation can lead to diagnostic errors (*e.g.*, the failure to diagnose sickle-cell disease in a European individual or cystic fibrosis in an Asian individual) or to inappropriate treatment or drug prescription. The general public, including policy-makers, are easily seduced by typological thinking, and so they must be made aware of the genetic data that help to prove it wrong.

A particular area of concern is in the genetics of human behavior. As genes that may influence behavior are identified, allele frequencies are often compared in populations[67,68]. These comparisons can produce useful evolutionary insights but can also lead to simplistic interpretations that may reinforce unfounded stereotypes[69]. In assessing the role of genes in population differences in behavior (real or imagined), several simple facts must be brought to the fore. Human behavior is complicated, and it is strongly influenced by nongenetic factors[70]. Thousands of pleiotropic genes are thought to influence behavior, and their products interact in complex and unpredictable ways. Considering this extraordinary complexity, the idea that variation in the frequency of a single allele could explain substantial population differences in behavior would be amusing if it were not so dangerous.

Race remains an inflammatory issue, both socially and scientifically. Fortunately, modern human genetics can deliver the salutary message that human populations share most of their genetic variation and that there is no scientific support for the concept that human populations are discrete, nonoverlapping entities. Furthermore, by offering the means to assess disease-related variation at the individual level, new genetic technologies may eventually render race largely irrelevant in the clinical setting. Thus, genetics can and should be an important tool in helping to both illuminate and defuse the race issue.

1. Smedley, A. Race in North America: Origin and Evolution of a Worldview (Westview, Boulder, 1999).
2. Sankar, P. & Cho, M.K. Genetics. Toward a new vocabulary of human genetic variation. *Science* **298**, 1337–1338 (2002).
3. Foster, M.W. & Sharp, R.R. Race, ethnicity, and genomics: social classifications as proxies of biological heterogeneity. *Genome Res.* **12**, 844–850 (2002).
4. Schwartz, R.S. Racial profiling in medical research. *N. Engl. J. Med.* **344**, 1392–1393 (2001).
5. Haga, S.B. & Venter, J.C. Genetics. FDA races in wrong direction. *Science* **301**, 466 (2003).

6. Risch, N., Burchard, E., Ziv, E. & Tang, H. Categorization of humans in biomedical research: genes, race and disease. *Genome Biol.* **3**, comment2007 (2002).

7. Burchard, E.G. *et al.* The importance of race and ethnic background in biomedical research and clinical practice. *N. Engl. J. Med.* **348**, 1170–1175 (2003).

8. Wood, A.J. Racial differences in the response to drugs—pointers to genetic differences. *N. Engl. J. Med.* **344**, 1394–1396 (2001).

9. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).

10. Schneider, J.A. *et al.* DNA variability of human genes. *Mech. Ageing Dev* **124**, 17–25 (2003).

11. Fischer, A., Wiebe, V., Paabo, S. & Przeworski, M. Evidence for a complex demographic history of chimpanzees. *Mol. Biol. Evol.* **21**, 799–808 (2004).

12. Li, W.-H. & Sadler, L.A. Low nucleotide diversity in man. *Genetics* **129**, 513–523 (1991).

13. Harpending, H.C. *et al.* Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. USA* **95**, 1961–1967 (1998).

14. Sakate, R. *et al.* Analysis of 5′-end sequences of chimpanzee cDNAs. *Genome Res.* **13**, 1022–1026 (2003).

15. Wildman, D.E. A map of the common chimpanzee genome. *Bioessays* **24**, 490–493 (2002).

16. Nei, M. & Livshits, G. Genetic relationships of Europeans, Asians and Africans and the origin of modern Homo sapiens. *Hum. Hered.* **39**, 276–281 (1989).

17. Cavalli-Sforza, L.L. & Bodmer, W.F. *The Genetics of Human Populations*, (W.H. Freeman, San Francisco, 1971).

18. Jorde, L.B. *et al.* The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y chromosome data. *Am. J. Hum. Genet.* **66**, 979–988 (2000).

19. Barbujani, G., Magagni, A., Minch, E. & Cavalli-Sforza, L.L. An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci. USA* **94**, 4516–4519 (1997).

20. Lewontin, R.C. The apportionment of human diversity. *Evol. Biol.* **6**, 381–398 (1972).

21. Watkins, W.S. *et al.* Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. *Genome Res.* **13**, 1607–1618 (2003).

22. Yu, N. *et al.* Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* **161**, 269–274 (2002).

23. Tishkoff, S.A. & Verrelli, B.C. Patterns of Human Genetic Diversity: Implications for Human Evolutionary History and Disease. *Annu. Rev. Genomics Hum. Genet.* **4**, 293–340 (2003).

24. Jorde, L.B., Watkins, W.S. & Bamshad, M.J. Population genomics: a bridge from evolutionary history to genetic medicine. *Hum. Mol. Genet.* **10**, 2199–2207 (2001).

25. Underhill, P.A. *et al.* Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**, 358–361 (2000).

26. Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713 (2000).

27. Calafell, F., Shuster, A., Speed, W.C., Kidd, J.R. & Kidd, K.K. Short tandem repeat polymorphism evolution in humans. *Eur. J. Hum. Genet.* **6**, 38–49 (1998).

28. Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton University Press, Princeton, 1994).

29. Jorde, L.B., Bamshad, M. & Rogers, A.R. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* **20**, 126–136 (1998).

30. Romualdi, C. *et al.* Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res.* **12**, 602–612 (2002).

31. Wilson, J.F. *et al.* Population genetic structure of variable drug response. *Nat. Genet.* **29**, 265–269 (2001).

32. Bamshad, M.J. *et al.* Human population genetic structure and inference of group membership. *Am. J. Hum. Genet.* **72**, 578–589 (2003).

33. Rosenberg, N.A. *et al.* Genetic structure of human populations. *Science* **298**, 2381–2385 (2002).

34. Shriver, M.D. *et al.* The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum. Genomics* **1**, 274–286 (2004).

35. Bowcock, A.M. *et al.* High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–457 (1994).

36. Mountain, J.L. & Cavalli-Sforza, L.L. Multilocus genotypes, a tree of individuals, and human evolutionary history. *Am. J. Hum. Genet.* **61**, 705–718 (1997).

37. Tishkoff, X. & Kidd, X. Title. *Nat. Genet.* **36** Supplement, '–' (2004). **[this issue]**

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

39. Edwards, A.W.F. Human genetic diversity: Lewontin's fallacy. *Bioessays* **25**, 798–801 (2003).

40. Bamshad, M. *et al.* Genetic evidence on the origins of Indian caste populations. *Genome Res.* **11**, 994–1104 (2001).

41. Feldman, M.W., Lewontin, R.C. & King, M.-C. A genetic melting-pot. *Nature* **424**, 374 (2003).

42. Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. Demic expansions and human evolution. *Science* **259**, 639–646 (1993).

43. Blumenbach, J.F. *On the Natural Varieties of Mankind (De Generis Humani Varietate Nativa).* Translated and edited by T. Bendyshe. (Bergman, New York, 1969).

44. Darwin, C. The Descent of Man, and Selection in Relation to Sex (John Murray, London, 1871).

45. Boyd, W.C. Genetics and the Races of Man: An Introduction to Physical Anthropology (Little Brown, Boston, 1950).

46. Exner, D.V., Dries, D.L., Domanski, M.J. & Cohn, J.N. Lesser response to angiotensin-converting-enzyme inhibitor therapy in black as compared with white patients with left ventricular dysfunction. *N. Engl. J. Med.* **344**, 1351–1357 (2001).

47. Varner, R.V., Ruiz, P. & Small, D.R. Black and white patients response to antidepressant treatment for major depression. *Psychiatr. Q.* **69**, 117–125 (1998).

48. Strickland, T.L., Stein, R., Lin, K.M., Risby, E. & Fong, R. The pharmacologic treatment of anxiety and depression in African Americans. Considerations for the general practitioner. *Arch. Fam. Med.* **6**, 371–375 (1997).

49. Bamshad, M., Wooding, S., Salisbury, B.A. & Stephens, J.C. Deconstructing the relationship between genetics and race. *Nat. Rev. Genet.* **5**, 598–609 (2004).

50. Evans, W.E. & Relling, M.V. Moving towards individualized medicine with pharmacogenomics. *Nature* **429**, 464–468 (2004).

51. Patil, N. *et al.* Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**, 1719–1723 (2001).

52. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).

53. Kunz, R., Kreutz, R., Beige, J., Distler, A. & Sharma, A.M. Association between the angiotensinogen 235T-variant and essential hypertension in whites: a systematic review and methodological appraisal. *Hypertension* **30**, 1331–1337 (1997).

54. Nakajima, T. *et al.* Natural selection and population history in the human angiotensinogen gene (AGT): 736 complete AGT sequences in chromosomes from around the world. *Am. J. Hum. Genet.* **74**, 898–916 (2004).

55. Weinshilboum, R. Inheritance and drug response. *N. Engl. J. Med.* **348**, 529–537 (2003).

56. Bradford, L.D. CYP2D6 allele frequency in European Caucasians, Asians, Africans and their descendants. *Pharmacogenomics* **3**, 229–243 (2002).

57. Cooper, R.S., Kaufman, J.S. & Ward, R. Race and genomics. *N. Engl. J. Med.* **348**, 1166–1170 (2003).

58. Sehgal, A.R. Overlap between whites and blacks in response to antihypertensive drugs. *Hypertension* **43**, 566–572 (2004).

59. Roses, A.D. Pharmacogenetics and the practice of medicine. *Nature* **405**, 857–865 (2000).

60. Cooper, R.S., Rotimi, C.N. & Ward, R. The puzzle of hypertension in African-Americans. *Sci. Am.* **280**, 56–63 (1999).

61. Laitinen, T. *et al.* Characterization of a common susceptibility locus for asthma-related traits. *Science* **304**, 300–304 (2004).

62. Altshuler, D. *et al.* The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26**, 76–80 (2000).

63. Hugot, J.P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).

64. Van Eerdewegh, P. *et al.* Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* **418**, 426–430 (2002).

65. Selkoe, D.J. & Podlisny, M.B. Deciphering the genetic basis of Alzheimer's disease. *Annu. Rev. Genomics Hum. Genet.* **3**, 67–99 (2002).

66. Chung, C.H., Bernard, P.S. & Perou, C.M. Molecular portraits and the family tree of cancer. *Nat. Genet.* **32** Suppl, 533–540 (2002).

67. Gilad, Y., Rosenberg, S., Przeworski, M., Lancet, D. & Skorecki, K. Evidence for positive selection and population structure at the human MAO-A gene. *Proc. Natl. Acad. Sci. USA* **99**, 862–867 (2002).

68. Ding, Y.C. *et al.* Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proc. Natl. Acad. Sci. USA* **99**, 309–314 (2002).

69. Harpending, H. & Cochran, G. In our genes. *Proc. Natl. Acad. Sci. USA* **99**, 10–12 (2002).

70. Bouchard, T.J. Jr. & Loehlin, J.C. Genes, evolution, and personality. *Behav. Genet.* **31**, 243–273 (2001).

71. Excoffier, L. & Hamilton, G. Comment on "Genetic structure of human populations". *Science* **300**, 1877 (2003).

72. Jorde, L.B. *et al.* Microsatellite diversity and the demographic history of modern humans. *Proc. Natl. Acad. Sci. USA* **94**, 3100–3103 (1997).

73. Watkins, W.S. *et al.* Patterns of ancestral human diversity: an analysis of *Alu* insertion and restriction site polymorphisms. *Am. J. Hum. Genet.* **68**, 738–752 (2001).

74. Felsenstein, J. PHYLIP (Phylogeny Inference Package) version 3.5c. (Department of Genetics, University of Washington, 1993).