

**The New York Times**

This copy is for your personal, noncommercial use only. You can order presentation-ready copies for distribution to your colleagues, clients or customers, please click here or use the "Reprints" tool that appears next to any article. Visit [www.nytreprints.com](http://www.nytreprints.com) for samples and additional information. Order a reprint of this article now. »

**BRENDAN  
GLEESON**

March 14, 2005

## A Family Tree in Every Gene

By Armand Marie Leroi

SHORTLY after last year's tsunami devastated the lands on the Indian Ocean, The Times of India ran an article with this headline: "Tsunami May Have Rendered Threatened Tribes Extinct." The tribes in question were the Onge, Jarawa, Great Andamanese and Sentinelese -- all living on the Andaman Islands -- and they numbered some 400 people in all. The article, noting that several of the archipelago's islands were low-lying, in the direct path of the wave, and that casualties were expected to be high, said, "Some beads may have just gone missing from the Emerald Necklace of India."

The metaphor is as colorful as it is well intentioned. But what exactly does it mean? After all, in a catastrophe that cost more than 150,000 lives, why should the survival of a few hundred tribal people have any special claim on our attention? There are several possible answers to this question. The people of the Andamans have a unique way of life. True, their material culture does not extend beyond a few simple tools, and their visual art is confined to a few geometrical motifs, but they are hunter-gatherers and so a rarity in the modern world. Linguists, too, find them interesting since they collectively speak three languages seemingly unrelated to any others. But the Times of India took a slightly different tack. These tribes are special, it said, because they are of "Negrito racial stocks" that are "remnants of the oldest human populations of Asia and Australia."

It's an old-fashioned, even Victorian, sentiment. Who speaks of "racial stocks" anymore? After all, to do so would be to speak of something that many scientists and scholars say does not exist. If modern anthropologists mention the concept of race, it is invariably only to warn against and dismiss it. Likewise many geneticists. "Race is social concept, not a scientific one," according to Dr. Craig Venter -- and he should know, since he was first to sequence the human genome. The idea that human races are only social constructs has been the consensus for at least 30 years.

But now, perhaps, that is about to change. Last fall, the prestigious journal *Nature Genetics* devoted a large supplement to the question of whether human races exist and, if so, what they mean. The journal did this in part because various American health agencies are making race

an important part of their policies to best protect the public -- often over the protests of scientists. In the supplement, some two dozen geneticists offered their views. Beneath the jargon, cautious phrases and academic courtesies, one thing was clear: the consensus about social constructs was unraveling. Some even argued that, looked at the right way, genetic data show that races clearly do exist.

The dominance of the social construct theory can be traced to a 1972 article by Dr. Richard Lewontin, a Harvard geneticist, who wrote that most human genetic variation can be found within any given "race." If one looked at genes rather than faces, he claimed, the difference between an African and a European would be scarcely greater than the difference between any two Europeans. A few years later he wrote that the continued popularity of race as an idea was an "indication of the power of socioeconomically based ideology over the supposed objectivity of knowledge." Most scientists are thoughtful, liberal-minded and socially aware people. It was just what they wanted to hear.

Three decades later, it seems that Dr. Lewontin's facts were correct, and have been abundantly confirmed by ever better techniques of detecting genetic variety. His reasoning, however, was wrong. His error was an elementary one, but such was the appeal of his argument that it was only a couple of years ago that a Cambridge University statistician, A.W.F. Edwards, put his finger on it.

The error is easily illustrated. If one were asked to judge the ancestry of 100 New Yorkers, one could look at the color of their skin. That would do much to single out the Europeans, but little to distinguish the Senegalese from the Solomon Islanders. The same is true for any other feature of our bodies. The shapes of our eyes, noses and skulls; the color of our eyes and our hair; the heaviness, height and hairiness of our bodies are all, individually, poor guides to ancestry.

But this is not true when the features are taken together. Certain skin colors tend to go with certain kinds of eyes, noses, skulls and bodies. When we glance at a stranger's face we use those associations to infer what continent, or even what country, he or his ancestors came from -- and we usually get it right. To put it more abstractly, human physical variation is correlated; and correlations contain information.

Genetic variants that aren't written on our faces, but that can be detected only in the genome, show similar correlations. It is these correlations that Dr. Lewontin seems to have ignored. In essence, he looked at one gene at a time and failed to see races. But if many -- a few hundred -- variable genes are considered simultaneously, then it is very easy to do so. Indeed, a 2002 study by scientists at the University of Southern California and Stanford showed that if a sample of people from around the world are sorted by computer into five groups on the basis of genetic similarity, the groups that emerge are native to Europe, East Asia, Africa, America and

Australasia -- more or less the major races of traditional anthropology.

One of the minor pleasures of this discovery is a new kind of genealogy. Today it is easy to find out where your ancestors came from -- or even when they came, as with so many of us, from several different places. If you want to know what fraction of your genes are African, European or East Asian, all it takes is a mouth swab, a postage stamp and \$400 -- though prices will certainly fall.

Yet there is nothing very fundamental about the concept of the major continental races; they're just the easiest way to divide things up. Study enough genes in enough people and one could sort the world's population into 10, 100, perhaps 1,000 groups, each located somewhere on the map. This has not yet been done with any precision, but it will be. Soon it may be possible to identify your ancestors not merely as African or European, but Ibo or Yoruba, perhaps even Celt or Castilian, or all of the above.

The identification of racial origins is not a search for purity. The human species is irredeemably promiscuous. We have always seduced or coerced our neighbors even when they have a foreign look about them and we don't understand a word. If Hispanics, for example, are composed of a recent and evolving blend of European, American Indian and African genes, then the Uighurs of Central Asia can be seen as a 3,000-year-old mix of West European and East Asian genes. Even homogenous groups like native Swedes bear the genetic imprint of successive nameless migrations.

Some critics believe that these ambiguities render the very notion of race worthless. I disagree. The physical topography of our world cannot be accurately described in words. To navigate it, you need a map with elevations, contour lines and reference grids. But it is hard to talk in numbers, and so we give the world's more prominent features -- the mountain ranges and plateaus and plains -- names. We do so despite the inherent ambiguity of words. The Pennines of northern England are about one-tenth as high and long as the Himalayas, yet both are intelligibly described as mountain ranges.

So, too, it is with the genetic topography of our species. The billion or so of the world's people of largely European descent have a set of genetic variants in common that are collectively rare in everyone else; they are a race. At a smaller scale, three million Basques do as well; so they are a race as well. Race is merely a shorthand that enables us to speak sensibly, though with no great precision, about genetic rather than cultural or political differences.

But it is a shorthand that seems to be needed. One of the more painful spectacles of modern science is that of human geneticists piously disavowing the existence of races even as they investigate the genetic relationships between "ethnic groups." Given the problematic, even vicious, history of the word "race," the use of euphemisms is understandable. But it hardly aids understanding, for the term "ethnic group" conflates all the possible ways in which people

differ from each other.

Indeed, the recognition that races are real should have several benefits. To begin with, it would remove the disjunction in which the government and public alike defiantly embrace categories that many, perhaps most, scholars and scientists say do not exist.

Second, the recognition of race may improve medical care. Different races are prone to different diseases. The risk that an African-American man will be afflicted with hypertensive heart disease or prostate cancer is nearly three times greater than that for a European-American man. On the other hand, the former's risk of multiple sclerosis is only half as great. Such differences could be due to socioeconomic factors. Even so, geneticists have started searching for racial differences in the frequencies of genetic variants that cause diseases. They seem to be finding them.

Race can also affect treatment. African-Americans respond poorly to some of the main drugs used to treat heart conditions -- notably beta blockers and angiotensin-converting enzyme inhibitors. Pharmaceutical corporations are paying attention. Many new drugs now come labeled with warnings that they may not work in some ethnic or racial groups. Here, as so often, the mere prospect of litigation has concentrated minds.

Such differences are, of course, just differences in average. Everyone agrees that race is a crude way of predicting who gets some disease or responds to some treatment. Ideally, we would all have our genomes sequenced before swallowing so much as an aspirin. Yet until that is technically feasible, we can expect racial classifications to play an increasing part in health care.

The argument for the importance of race, however, does not rest purely on utilitarian grounds. There is also an aesthetic factor. We are a physically variable species. Yet for all the triumphs of modern genetics, we know next to nothing about what makes us so. We do not know why some people have prominent rather than flat noses, round rather than pointed skulls, wide rather than narrow faces, straight rather than curly hair. We do not know what makes blue eyes blue.

One way to find out would be to study people of mixed race ancestry. In part, this is because racial differences in looks are the most striking that we see. But there is also a more subtle technical reason. When geneticists map genes, they rely on the fact that they can follow our ancestors' chromosomes as they get passed from one generation to the next, dividing and mixing in unpredictable combinations. That, it turns out, is much easier to do in people whose ancestors came from very different places.

The technique is called admixture mapping. Developed to find the genes responsible for racial differences in inherited disease, it is only just moving from theory to application. But through it, we may be able to write the genetic recipe for the fair hair of a Norwegian, the black-verging-on-purple skin of a Solomon Islander, the flat face of an Inuit, and the curved eyelid of a Han

Chinese. We shall no longer gawp ignorantly at the gallery; we shall be able to name the painters.

There is a final reason race matters. It gives us reason -- if there were not reason enough already -- to value and protect some of the world's most obscure and marginalized people. When the Times of India article referred to the Andaman Islanders as being of ancient Negrito racial stock, the terminology was correct. Negrito is the name given by anthropologists to a people who once lived throughout Southeast Asia. They are very small, very dark, and have peppercorn hair. They look like African pygmies who have wandered away from Congo's jungles to take up life on a tropical isle. But they are not.

The latest genetic data suggest that the Negritos are descended from the first modern humans to have invaded Asia, some 100,000 years ago. In time they were overrun or absorbed by waves of Neolithic agriculturalists, and later nearly wiped out by British, Spanish and Indian colonialists. Now they are confined to the Malay Peninsula, a few islands in the Philippines and the Andamans.

Happily, most of the Andamans' Negritos seem to have survived December's tsunami. The fate of one tribe, the Sentinelese, remains uncertain, but an Indian coast guard helicopter sent to check up on them came under bow and arrow attack, which is heartening. Even so, Negrito populations, wherever they are, are so small, isolated and impoverished that it seems certain that they will eventually disappear.

Yet even after they have gone, the genetic variants that defined the Negritos will remain, albeit scattered, in the people who inhabit the littoral of the Bay of Bengal and the South China Sea. They will remain visible in the unusually dark skin of some Indonesians, the unusually curly hair of some Sri Lankans, the unusually slight frames of some Filipinos. But the unique combination of genes that makes the Negritos so distinctive, and that took tens of thousands of years to evolve, will have disappeared. A human race will have gone extinct, and the human species will be the poorer for it.

Drawing (Drawing by Christophe Vorlet)



---

## Confusions About Human Races

By R.C. Lewontin

**Published on: Jun 07, 2006**

*R.C. Lewontin, Alexander Agassiz Professor Emeritus of Zoology at Harvard University, has written a number of books and articles on evolution and human variation, including *Biology as Ideology: The Doctrine of DNA and The Triple Helix: Gene, Organism, and Environment**

Over the last thirty five years a major change has taken place in our biological understanding of the concept of human “race,” largely as a consequence of an immense increase in our knowledge of human genetics. As a biological rather than a social construct, “race” has ceased to be seen as a fundamental reality characterizing the human species. Nevertheless, there appear from time to time claims that racial categories represent not arbitrary socially and historically defined groups but objective biological divisions based on genetic differences. The most recent widely noticed rebirth of such claims is an essay by Armand Marie Leroi on the Op-Ed page of *The New York Times* (March 14, 2005), an essay that illustrates both the classical confusions about the reality of racial categories and the more recent erroneous conclusions about the relevance of such racial identifications for medical practice.

There are four facts about human variation upon which there is universal agreement. First, the human species as a whole has immense genetic variation from individual to individual. Any two unrelated human beings differ by about 3 million distinct DNA variants.

Second, by far the largest amount of that variation, about 85%, is among individuals within local national or linguistic populations, within the French, within the Kikuyu, within the Japanese. There is diversity from population to population in how much genetic variation each contains, depending upon how much immigration into the population has occurred from a variety of other groups and also on the size of the population. The United States, with a very large population whose ancestors came from all over the earth including the original inhabitants of the New World, is genetically very variable whereas small populations of local Amazonian tribes are less genetically variable, although they are by no means genetically uniform. Despite the differences in amount of genetic variation within local populations, the finding that on the average 85% of all human genetic variation is within local populations has been a remarkably consistent result of independent studies carried out over twenty-five years using data from both proteins and DNA.

Of the remaining 15% of human variation, between a quarter and a half is between local populations

within classically defined human "races," between the French and the Ukrainians, between the Kikuyu and the Ewe, between the Japanese and the Koreans. The remaining variation, about 6% to 10% of the total human variation is between the classically defined geographical races that we think of in an everyday sense as identified by skin color, hair form, and nose shape. This imprecision in assigning the proportion of variation assigned to differences among population within "races" as compared to variation among "races," arises precisely because there is no objective way to assign the various human populations to clear-cut races. Into which "race" do the Hindi and Urdu speakers of the Indian sub-continent fall? Should they be grouped with Europeans or with Asians or should a separate race be assigned to them? Are the Lapps of Finland and the Hazari of Afghanistan really Europeans or Asians? What about Indonesians and Melanesians? Different biologists have made different assignments and the number of "races" assigned by anthropologists and geneticists has varied from 3 to 30.

Third, a small number of genetic traits, such as skin color, hair form, nose shape (traits for which the genes have not actually been identified) and a relatively few proteins like the Rh blood type, vary together so that many populations with very dark skin color will also have dark tightly curled hair, broad noses and a high frequency of the Rh blood type Ro. Those who, like Leroi, argue for the objective reality of racial divisions claim that when such covariation is taken into account, clear-cut racial divisions will appear and that these divisions will correspond largely to the classical division of the world into Whites, Blacks, Yellows, Reds and Browns. It is indeed possible to combine the information from covarying traits into weighted averages that take account of the traits' covariation (technically known as "principal components" of variation). When this has been done, however, the results have not borne out the claims for racial divisions. The geographical maps of principal component values constructed by Cavalli, Menozzi and Piazza in their famous *The History and Geography of Human Genes* show continuous variation over the whole world with no sharp boundaries and with no greater similarity occurring between Western and Eastern Europeans than between Europeans and Africans! Thus, the classically defined races do not appear from an unprejudiced description of human variation. Only the Australian Aborigines appear as a unique group.

A clustering of populations that does correspond to classical continental "races" can be achieved by using a special class of non-functional DNA, microsatellites. By selecting among microsatellites, it is possible to find a set that will cluster together African populations, European populations, and Asian populations, etc. These selected microsatellite DNA markers are not typical of genes, however, but have been chosen precisely because they are "maximally informative" about group differences. Thus, they tell us what we already knew about the differences between populations of the classical "races" from skin color, face shape, and hair form. They have the added advantage of allowing us to make good estimates of the amount of intermixture that has occurred between populations as a result of migrations and conquests.

The every-day socially defined geographical races do identify groups of populations that are somewhat more closely similar to each other genetically. Most important from the standpoint of the biological meaning of these racial categories, however, most human genetic variation does not show such "race" clustering. For the vast majority of human genetic variations, classical racial categories as defined by a combination of geography, skin color, nose and hair shape, an occasional blood type or selected microsatellites make no useful prediction of genetic differences. This failure of the clustering of local populations into biologically meaningful "races" based on a few clear genetic differences is not confined to the human species. Zoologists long ago gave up the category of "race" for dividing up groups of animal populations within a species, because so many of these races turned out to be based on only one or two genes so that two animals born in the same litter could belong to different "races."

In his article, Leroi is inconsistent and shifting in his notion of race. Sometimes it corresponds to the classical social definitions of major races, but elsewhere he makes "race" coincident with a small local group such as the Negritos or Inuit. In this shifting concept of "race" he goes back to the varying use of the term in the 19th century. Then people spoke of the "Scots race," "the Irish race" and the "race of Englishmen." Indeed "race" could stand for a family group defined by male inheritance, as in the description of the last male in a family line as "the last of his race." This inconsistent usage arises from the fact that there is no clear criterion of how much difference between groups of genetically related individuals should correspond to the category "race." If it had turned out that groups of related populations were clearly different in the great majority of their genes from other groups, then racial categories would be clear and unambiguous and they would have great predictive power for as yet unstudied characters. But that is not the way it has turned out, at least for the human species.

The fourth and last fact about genetic differences between groups is that these differences are in the process of breaking down because of the very large amount of migration and intergroup mating that was always true episodically in the history of the human species but is now more widespread than ever. The result is that individuals identified by themselves or others as belonging to one "race," based on the small number of visible characters used in classical race definitions, are likely to have ancestry that is a mixture of these groups, a fact that has considerable significance for the medical uses of race identification.

A common claim, repeated by Leroi, is that racial categories are of considerable medical use, especially in diagnostic testing because some genetic disorders are very common in ancestral racial populations. For example sickle cell anemia is common among West Africans, who were brought as slaves to the New World, and Tay-Sachs disease is common among Ashkenazi Jews. So, it is argued, racial information can be a useful diagnostic indicator. Certainly classical "race" contains some medically relevant information in some cases, as for example "white" as opposed to "African American" if the contrast is between Finland and West Africa, but not if it is a contrast between a



“white” Mediterranean and an “Asian” Indian. There is a confusion here between race and ancestry. Sickle cell anemia is in high frequency not only in West Africans but also in some “white” Middle Eastern and Indian populations. Moreover, a person with, say, one African great-grandparent, but who is identified by herself and others as “white” has a one in eight chance of inheriting a sickle-cell mutation carried by that ancestor. There are, in addition, a number of other simply inherited hemoglobin abnormalities, the thalassemys, that are in high frequency in some places in the Mediterranean (Sardinia), Arabia and southeast Asia. The highest frequency known for a thalassemy (80%) is in Nepal, but it is rare in most of Asia. The categorization of individuals simply as “white” or “Afro-American” or “Asian” will result in a failure to test for such abnormal hemoglobins because these abnormalities do not characterize the identified “race” of the patient. Even group identities below the level of the conventional races are misleading. Two of my incontrovertibly WASP grandchildren have a single Ashkenazi Jewish great-grandparent and so have a one in eight chance of inheriting a Tay-Sachs abnormality carried by that ancestor. For purposes of medical testing we do not want to know whether a person is “Hispanic” but rather whether that person’s family came from a Caribbean country such as Cuba, that had a large influx of West African slaves, or one in which there was a great deal of intermixture with native American tribes as in Chile and Mexico, or one in which there was only a negligible population of non-Europeans. Racial identification simply does not do the work needed. What we ought to ask on medical questionnaires is not racial identification, but ancestry. “Do you know of any ancestors who were (Ashkenazi Jews, or from West Africa, from certain regions of the Mediterranean, from Japan)?” Once again, racial categorization is a bad predictor of biology.

There has been an interesting dialectic between the notion of human races and the use of race as a general biological category. Historically, the concept of race was imported into biology, and not only the biology of the human species, from social practice. The consciousness that human beings come in distinct varieties led, in the history of biology, to the construction of “race” as a subgrouping within species. For a long time the category “race” was a standard taxonomic level. But the use of “race” in a general biological context then reinforced its application to humans. After all, lots of animal and plant species are divided into races, so why not *Homo sapiens*? Yet the classification of animal and plant species into named races was at all times an ill-defined and idiosyncratic practice. There was no clear criterion of what constituted a race of animals or plants that could be applied over species in general. The growing realization in the middle of the twentieth century that most species had some genetic differentiation from local population to local population led finally to the abandonment in biology of any hope that a uniform criterion of race could be constructed. Yet biologists were loathe to abandon the idea of race entirely. In an attempt to hold on to the concept while make it objective and generalizable, Th. Dobzhansky, the leading biologist in the study of the genetics of natural populations, introduced the “geographical race,” which he defined as any population that differed genetically in any way from any other population of the species. But as genetics developed and it became possible to characterize the genetic differences between individuals and populations it became apparent, that *every* population of every species in fact

differs genetically to some degree from every other population. Thus, every population is a separate “geographic race” and it was realized that nothing was added by the racial category. The consequence of this realization was the abandonment of “race” as a biological category during the last quarter of the twentieth century, an abandonment that spread into anthropology and human biology. However, that abandonment was never complete in the case of the human species. There has been a constant pressure from social and political practice and the coincidence of racial, cultural and social class divisions reinforcing the social reality of race, to maintain “race” as a human classification. If it were admitted that the category of “race” is a purely social construct, however, it would have a weakened legitimacy. Thus, there have been repeated attempts to reassert the objective biological reality of human racial categories despite the evidence to the contrary.