Wade, Nicholas. A Troublesome Inheritance: Genes, Race, and Human History. New York: Penguin Press, 2014.

5

THE GENETICS OF RACE

> Selfish and contentious people will not cohere, and without coherence nothing can be effected. A tribe rich in the above qualities would spread and be victorious over other tribes: but in the course of time it would, judging from all past history, be in its turn overcome by some other tribe more highly endowed. Thus the social and moral qualities would tend slowly to advance and be diffused throughout the world.
>
> —CHARLES DARWIN[1]

In the case of human races, the genetic differences from one race to another are slight and subtle. One might expect that different races would have different genes, but they don't. All humans, so far as is known, have the same set of genes. Each gene comes in various alternative forms, called alleles, so the next expectation might be that races would be distinguished by having different alleles of various genes. But this too is not how the system works. There are a mere handful of known cases where a particular allele of a gene occurs in only one race.

The genetic differences between human races turn out to be based largely in allele frequencies, meaning the percentages of each allele that occur in a given race. How a mere difference in allele frequencies could lead to differences in physical traits is explained below.

Races as Clusters of Variation

A useful approach to studying racial variation is to look not for absolute differences but at how the genomes of individuals throughout the world cluster together in terms of their genetic similarity. The result is that everyone ends up in the cluster with which they share the most variation in common. These clusters always correspond to the five continental races in the first instance, though when extra DNA markers are used, the people of the Indian subcontinent sometimes split away from Caucasians as a sixth major group, and people of the Middle East as a seventh.

One of the first genetic clustering techniques depended on examining an element of the genome called tandem repeats. There are many sites on the genome where the same pair of DNA units is repeated several times in tandem. CA stands for the DNA unit known as a cytosine followed by adenine, so the DNA sequence CACACACA would be called a tandem CA repeat. The string of repeats occasionally confuses the DNA copying apparatus, which every few generations may add or drop a repeat unit during the copying process that has to occur before a cell can divide. Sites at which repeats occur therefore tend to be quite variable, and this variability is useful for comparing populations.

In 1994, in one of the earliest attempts to study human differentiation in terms of DNA differences, a research team led by Anne Bowcock of the University of Texas and Luca Cavalli-Sforza of Stanford University looked at CA repeats at 30 sites on the genome in people from 14 populations. Comparing their subjects on the basis of the number of CA repeats at each genomic site, the researchers found that people clustered together in groups that were coincident with their continent of origin. In other words, all the Africans had patterns of CA repeats that resembled one another, all the American Indians had a different pattern of repeats and so on. Altogether there were 5 principal clusters of CA repeats, formed by people living in each of the 5 continental regions of Africa, Europe, East Asia, the Americas and Australasia.[2]

Many larger and more sophisticated surveys have been done since, and all have come to the same conclusion, that “genetic differentiation is greatest when defined on a continental basis,” writes Neil Risch, a statistical geneticist at the University of California, San Francisco. “Effectively, these

population genetic studies have recapitulated the classical definition of races based on continental ancestry—namely African, Caucasian (Europe and Middle East), Asian, Pacific Islander (for example, Australian, New Guinean and Melanesian), and Native American."[3]

In one of these more sophisticated studies, a team led by Noah Rosenberg of the University of Southern California and Marcus Feldman of Stanford University looked at the number of repeats at 377 sites on the genome of more than 1,000 people around the world. When this many sites are examined on a genome, it's possible to assign segments of an individual's genome to different races if he or she has mixed ancestry. This is because each race or ethnicity has a characteristic number of repeats at each genomic site.

The Rosenberg-Feldman study showed, as expected, that the 1,000 individuals in their study clustered naturally into five groups, corresponding to the five continental races. Feldman, the senior author and tutor of many American population geneticists, said when the study was published that it essentially confirmed the popular conception of race and the statement by Neil Risch that genetics confirms the definition of race by continental ancestry. "Neil's article was theoretical and this is the data that backs up what he said," Feldman remarked.[4]

Other leading geneticists also regard the continent-based clustering of human variation as corresponding to the general notion of race. "There are difficulties in where you put boundaries on the globe, but we know now there are enough genetic differences between people from different parts of the world that you can classify people in groups that correspond to popular notions of race," said Jonathan Pritchard of the University of Chicago.[5]

The Rosenberg-Feldman study also brought out the fact that several Central Asian ethnicities, such as Pathans, Hazara and Uigurs, are of mixed European and East Asian ancestry. This is not a surprise, given the frequent movement of peoples to and fro across Central Asia.

Language is often an isolating mechanism that deters intermarriage with neighboring groups. The Burusho, a people of Pakistan who speak a unique language, turn out also to be unlike their neighbors genetically. Within races, the Rosenberg-Feldman study showed that different ethnicities could be recognized. Among Africans, it is easy to distinguish by their genomes the Yoruba of Nigeria, the San (a click-speaking people of southern Africa) and the Mbuti and Biaka pygmies.

Many populations are not highly mixed, and the Rosenberg-Feldman survey confirmed the remarkable extent to which people throughout history have lived and died in the place where they were born.[6]

In the ancestral human population in Africa, a large number of alleles had developed for each gene over many generations. Those who migrated out of Africa took away only a sample of these alleles. And each time a new group split off, the number of alleles from the original population again decreased.

The farther away from Africa that this process continued, the less was the diversity of alleles. This downhill gradient happens with any population that expands too far from its origins to maintain the regular interbreeding that keeps the gene pool well mixed.

A genetic gradient, or cline, is what some researchers prefer to think exists in place of races. “There are no races, there are only clines,” asserted the biological anthropologist Frank Livingstone.[7] Critics raised the same objection against the Rosenberg-Feldman result, alleging that the clustering of individuals into races was an artifact and that with a geographically more uniform sampling approach, the researchers would have seen only clines.[8] The Rosenberg-Feldman team then reanalyzed their data and gave their survey finer resolution by looking at 993 sites, not just 377, on each of the genomes in their study. They found that the clusters are real. Although there are gradients of genetic diversity, there is also a clustering into the continental groups described in their first article.[9]

Rosenberg and Feldman compared people’s genomes on the basis of DNA repeats. Another kind of DNA marker has since become available for global population comparison—the SNP, which is more useful for medical studies. SNP stands for single nucleotide polymorphism, meaning a site on the genome where some people have a different kind of DNA unit from that of the majority. A vast preponderance of sites on the genome are fixed, meaning everyone has the same DNA unit, whether A, T, G or C. The fixed sites, being all the same, say nothing about human variation. It’s the SNP sites, which are variable, that are of particular interest to geneticists because they afford a direct way of comparing populations. To exclude the many random mutations that occur just in particular individuals and have no wider importance, SNPs are arbitrarily defined as sites on the genome where at least 1% of the population has a DNA unit other than the standard one.

A research group led by Jun Z. Li and Richard M. Myers has applied a

clustering program like that used by Rosenberg and Feldman to almost 1,000 people in 51 populations across the globe. Each person's genome was examined at 650,000 SNP sites. On the basis of SNPs, just as with the DNA repeats, people sampled from around the world clustered into 5 continental groups. But in addition, the SNP library brought to light two other major clusters. These had not emerged in the Rosenberg-Feldman study, which had used fewer markers. The more DNA markers that are used, whether tandem repeats or SNPs, the more subdivisions can be established in the human population.

One of the new clusters is formed by the people of Central and South Asia, including India and Pakistan. The second is the Middle East, where there is considerable admixture between people from Europe and Africa.[10] It might be reasonable to elevate the Indian and Middle Eastern groups to the level of major races, making seven in all. But then many more subpopulations could be declared races, so to keep things simple, the five-race, continent-based scheme seems the most practical for most purposes.

Some readers may be troubled that the number of human races is not fixed but depends on the way race is assessed. But this should not be a surprise, given that races are not distinct entities but rather clusters of individuals with similar genetic variation. How many hills are there in New Hampshire? The answer depends on the height selected to qualify as a hill. The number of human races depends on the degree of clustering to be recognized, and three, five and seven are all reasonable answers to the issue of enumerating the major subsets of human variation.

Within each continental race, the SNP analysis could separate out further subgroups. Within Europe it distinguished French, Italians, Russians, Sardinians and Orcadians (people who live in the Orkney Islands, north of Scotland). In China the northern Han can be distinguished from the southern Han.

Groupings within Africa are of particular interest because this is where modern humans spent the first 150,000 years of their existence. In the most thorough survey of Africa so far, Sarah Tishkoff and colleagues surveyed people from 121 populations, scanning their genomes at 1,327 variable sites, most of them DNA repeats. The survey brought to light 14 different ancestral groups within Africa. Tishkoff found that, unlike in the rest of the world, where there are definable continental races, in Africa most populations are admixtures of several ancestral groups. There have presumably been a larger

number of migration events within Africa, which served to mix up populations that were originally separate. The most recent large-scale migration was the Bantu expansion, a population explosion driven by new agricultural technology. Within the past few thousand years, Bantu speakers from the region of Nigeria and Cameroon in West Africa have migrated across to eastern Africa and down both coasts to southern Africa. Only a few groups have kept relatively clear of the churning of populations within Africa. These include the click-speaking peoples of Tanzania and southern Africa, who until recently have been hunter-gatherers, and the various pygmy groups, who live deep in the forest.[11]

The click-speakers and pygmies may be remnants of a much earlier hunter-gatherer population that once occupied a large part of southern Africa and the eastern coast as far north as Somalia. The click-speakers speak a group of languages known as Khoisan, which are unlike any others and have only very distant relationships among themselves, probably reflecting their great antiquity. The pygmy groups too may once have spoken Khoisan languages but it is impossible to know for sure, because they have lost their original languages.

Africa has four language superfamilies, of which Khoisan is one and the other three are Niger-Kordofanian (also known as Niger-Congo), Nilo-Saharan and Afro-Asiatic. The Niger-Kordofanian languages, the most widespread, were carried from western to eastern Africa and then south by the Bantu expansion, a great stream of migrations from the proto-Bantu homeland in western Africa that began in about 1000 BC and reached southern Africa a thousand years later. Afro-Asiatic languages are spoken in a broad belt across northern Africa, and the Nilo-Saharan speakers are sandwiched between Afro-Asiatic to the north and Niger-Kordofanian to the south.

Genetics generally correlates with language family, except in the case of populations that have switched languages; the pygmies now speak Niger-Kordofanian languages, and the Luo of Kenya, whose genetics place them with Niger-Kordofanian speakers, now speak a Nilo-Saharan language.

The Tishkoff team surveyed African Americans from Chicago, Baltimore, Pittsburgh and North Carolina and found that 71% of their genomes, on average, matched the genetics of Niger-Kordofanian speakers, 8% matched that of other African populations and 13% were European. These percentages varied greatly from one individual to another.

The origin of a species can often be located by surveying the genetic diversity in its members and seeing where diversity is highest. This is because the founding population will have had longest to accumulate the mutations that generate diversity, and the groups that migrate away will carry with them only a sample of the original mutations. (Other forces, like natural selection, reduce diversity by eliminating harmful mutations and sweeping away others when a beneficial mutation is favored.) On the basis of the new African and other genomic data, the origin of the modern human migration lies in southwestern Africa, near the border of Namibia and Angola, in a region that is the current homeland of the San click-speakers. The finding is not definitive, because the distribution of ancient populations may have been rather different from those of today. Nonetheless, the fact that human genetics points to a single origin confirms that today's races are all mere variations on the same theme.

Fingerprints of Selection in the Human Genome

Both repeated DNA units and SNPs, the two kinds of DNA marker used by the surveys described above, lie for the most part outside genes and have little or no effect on a person's physical makeup. They are what geneticists call neutral variations, meaning that they are ignored by natural selection. What then is it that makes human populations differ from one another?

Natural selection is the major shaper of differences, especially in large societies. In small societies, genetic drift—the luck of the draw as to which alleles make it into the next generation—can be a significant influence. But natural selection, often in concert with drift, is a major force over the long run. With the advent of fast methods of genome sequencing, geneticists have at last begun to delineate the fingerprints of natural selection in remodeling the human genome. These fingerprints are both recent and regional, meaning that they differ from one race to another.

The regional nature of selection was first made evident in a genomewide

scan undertaken by Jonathan Pritchard, a population geneticist at the University of Chicago, in 2006. He looked for genes under selection in the three major races—Africans, East Asians and Europeans (or more exactly Caucasians, but European genetics are at present much better understood, so European populations are the usual subjects of study). Copious genetic data had been collected on each race as part of the HapMap, a project undertaken by the National Institutes of Health to explore the genetic roots of common disease. In each race Pritchard found about 200 genetic regions that showed a characteristic signature of having been under selection (206 in Africans, 185 in East Asians and 188 in Europeans). But in each race, a largely different set of genes was under selection, with only quite minor overlaps.[12]

The evidence of natural selection at work on a gene is that the percentage of the population that carries the favored allele of the gene has increased. But though alleles under selection become more common, they rarely displace all the other alleles of the gene in question by attaining a frequency of 100%. Were this to happen often in a population, races could be distinguished on the basis of which alleles they carried, which is generally not the case. In practice, the intensity of selection often relaxes as an allele rises in frequency, because the needed trait is well on the way to being attained.

Geneticists have several tests for whether a gene has been a recent target for natural selection. Many such tests, including the one devised by Pritchard, rest on the fact that as the favored allele of a gene sweeps through a population, the amount of genetic diversity in and around the gene is reduced in the population as a whole. This is so because increasing numbers of people now carry the same sequence of DNA units at that site, those of the favored allele. So the result of such a sweep is that DNA differences between members of a population are reduced in the region of the genome affected by the sweep. The concept of using sweeps as signatures of natural selection is discussed further below.

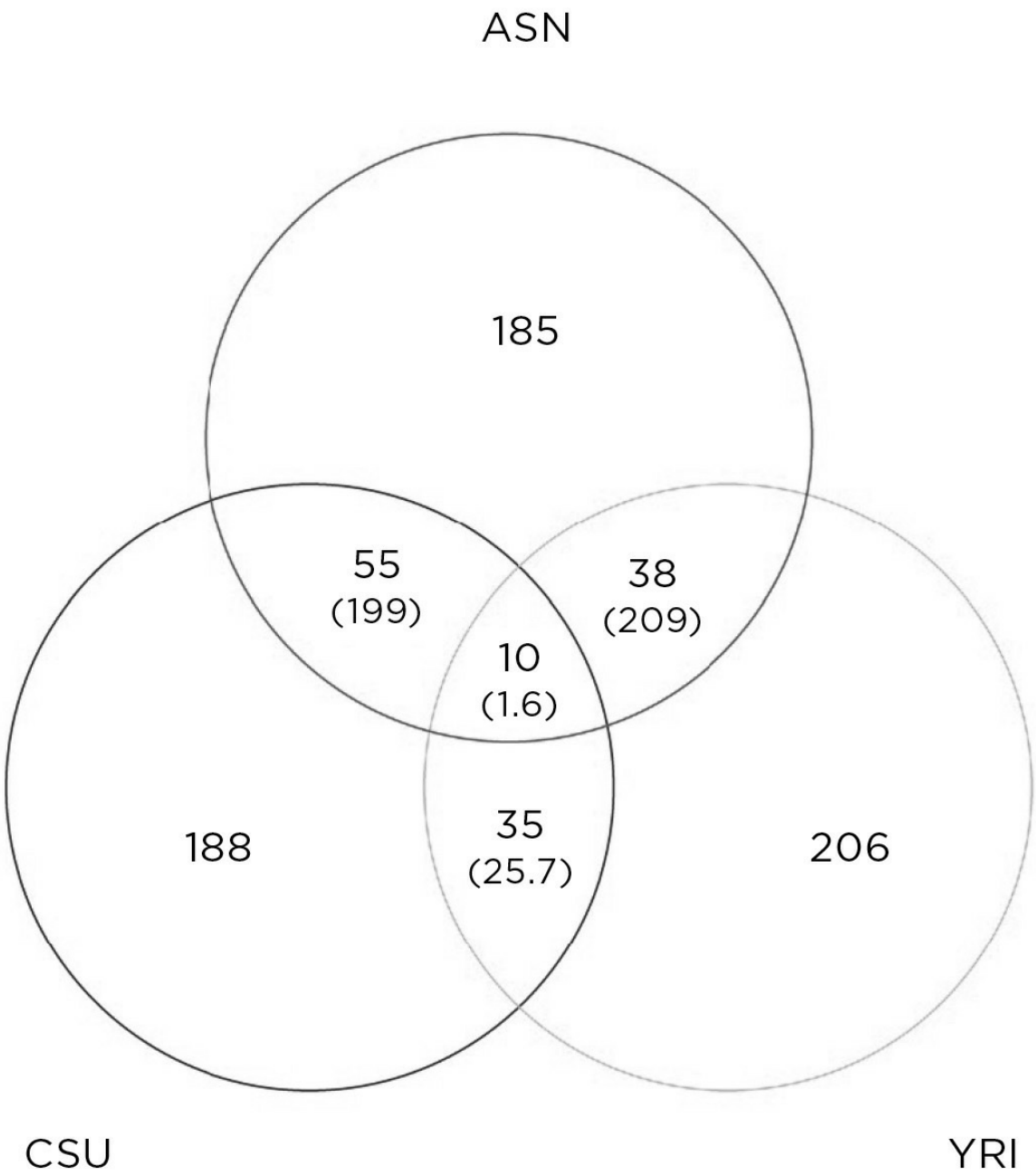

Figure 4.1. Regions of the genome that are highly selected in the three major races. ASN=East Asian, a sample of Chinese and Japanese. YRI=Yoruba, a West African people. CEU=European.

FROM JONATHAN PRITCHARD, *PLOS BIOLOGY* 4(2006):446–58.

Other researchers too have found that in doing genome scans for the fingerprints of natural selection, each major race or continental population has its own distinctive set of sites where selection has occurred.

These sites of selection are often very large and contain many genes,

making it hard or impossible to decide which specific gene was the target of natural selection. In a new approach, which takes advantage of the many whole genomes that have now been decoded, Pardis Sabeti of Harvard and colleagues have defined 412 regions under selection in Africans, Europeans and East Asians. The regions are so small that most contain one or no genes. Those without genes presumably contain a control element, meaning a stretch of DNA that regulates some nearby gene.[13]

Of the 412 regions of the human genome shown to be under selection, 140 were under selection just in Europeans, 140 in East Asians and 132 in Africans.[14] The absence of any overlap, meaning genes selected in two or more populations, as was found by Pritchard, is due to the Sabeti team's genome scanning method, which depended in part on looking for sites at which the three races differed.

Each gene under selection will eventually tell a fascinating story about some historical stress to which the population was exposed and then adapted. A case in point is the analysis of the EDAR-V370A allele which, as described in the previous chapter, is the cause of thick hair and other traits in East Asians. But those narratives are for the moment inaccessible. The exploration of the human genome is so much at its beginning that the precise function of most genes is unknown.

Still, even though the exact tasks of most genes are still uncertain, the general roles of most genes can be inferred by comparing the DNA sequence of any unknown gene with those of known genes recorded in genomic data banks. The known genes are grouped into general functional categories, like brain genes or genes involved in metabolism, and since function is related to structure, the genes in each category have a characteristic sequence of DNA units. By comparing the DNA sequence of any new gene with the data bank sequences, the gene can be assigned to a general functional category. The genes Pritchard identified as shaped by natural selection included genes for fertilization and reproduction, genes for skin color, genes for skeletal development and genes for brain function. In the brain function category, four genes were under selection in Africans and two each in East Asians and Europeans. What these genes do within the brain is largely unknown. But the findings establish the obvious truth that brain genes do not lie in some special category exempt from natural selection. They are as much under evolutionary pressure as any other category of gene.

Population geneticists have developed several different kinds of tests to

see if natural selection has influenced the DNA sequence of a gene. All these tests are statistical, and many depend on the disturbance in gene frequencies that is caused as a favored gene sweeps through a population. Natural selection cannot pick out single genes or even single mutations in DNA. Rather, it depends on the process called recombination, in which the mother's and father's genomes are shuffled prior to creating eggs and sperm.

In the egg-making or sperm-making cells, the two sets of chromosomes that a person has inherited, one from their mother and one from their father, are lined up side by side, and the cell then forces them to exchange large sections of DNA. These new composite chromosomes, consisting of some sections from the father's genome and some from the mother's, are what is passed on to the next generation.

The swapped sections, or blocks, may be 500,000 DNA units in length, long enough to carry several genes. So a gene with a beneficial mutation will be inherited along with the whole block of DNA in which it is embedded. It's because beneficial genes lie in such a large block that the effect of natural selection on the genome can be detected—the favored blocks sweep out large regions of the genome as they spread through a population.

Generation by generation, the block of DNA with the favored version of a gene gets to be carried by more and more people. Eventually, the new allele may sweep through the entire population, in which case geneticists say it has gone to fixation. But most sweeps do not carry an allele to fixation because, as already noted, the intensity of selection on a beneficial allele relaxes as the trait is molded toward its most efficient form.

Whether a sweep is complete or partial, the favored blocks of DNA eventually get whittled down over the generations, because the cuts that generate them are not always made in the same places on the chromosome. After just 30,000 years or so, according to one calculation, the blocks get too short to be detectable. This means that most genomewide scans for selection are looking at events that occurred just a few thousand years ago, very recently in human evolution.

Biologists have long had to depend on the evidence from fossils to judge the speed of evolution. But fossils capture just the bones of an animal. And since the skeletal structure of a species changes only slowly, evolution has long seemed a glacially slow and plodding process.

With the ability to decode DNA sequences, biologists can examine the raw programming of evolutionary change and track every gene in a species'

repertoire. It's now clear that evolution is no sluggard. There are already clear examples of human evolutionary change within the past few thousand years, such as the continued evolution of European skin, hair and eye color within the past 5,000 years. Of course, every gene in the human genome has been intensely shaped by natural selection at one time or another. But with most genes, the selection was accomplished eons before humans or even primates had evolved. The fingerprints of these ancient selection events have long since faded from sight. The type of selection picked up by most genome scans is very recent selection, meaning within the past 5,000 to 30,000 years or so, but fortunately this is a period of great interest for understanding human evolution.

More than 20 scans for selection have now been performed on the human genome. They do not all mark the same regions as being under selection but that is not surprising since the authors use different kinds of tests and different statistical methods, which are in any case imprecise. But if one takes just the regions marked by any two of the scans, then 722 regions, containing some 2,465 genes, have been under recent pressure of natural selection, according to an estimate by Joshua M. Akey of the University of Washington. This amounts to at least 8% of the genome.[15]

That so much of the genome has been under natural selection strong enough to be detectable shows how intense human evolution must have been in the past few thousand years. A principal driver of evolutionary change would have been the need to adapt to a wide range of new environments. In proof of that point, some 80% of the 722 regions under selection are instances of local adaptation, meaning that they occur in one of the three main races but not in the other two.

The genes under selection affect a large number of biological traits, prominent among them being skin color, diet, bone and hair structure, resistance to disease and brain function.

A similar finding emerged from a particularly comprehensive genome scan conducted by Mark Stoneking and colleagues. Stoneking, a population geneticist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, is known for having developed an ingenious way of estimating when humans first started to wear clothes. The body louse, which lives only in clothes, evolved from the head louse, which lives on hair. Stoneking realized that a date for the first tight-fitting clothes could be derived by using genetic methods to date the birth of the body louse lineage—about 72,000 years

ago.[16]

In his genome survey, Stoneking found many genes under selection that affected people's interaction with their environment, such as genes involved in metabolizing certain classes of food and genes that mediate resistance to pathogens. Among the genes under selection he also found several that were involved in aspects of the nervous system, such as cognition and sensory perception.

The genes of the nervous system have been under selection for the same reason as the other genes—to help people adapt to local circumstances. Changes in social behavior may well have been foremost, given that it is largely through their society that people interact with their environment. Signals of selection in brain genes "may be related to how different human groups interact behaviorally with their environment and/or with other human groups," Stoneking and colleagues wrote.[17]

Another regional trend indicated by the genome scans is that there seem to be more genes under selection in the genomes of East Asians and Europeans than in those of Africans. Not all genome scans have reported such a finding—the Pritchard scan described above did not—and African populations have been poorly sampled so far. But in a subsequent scan, Pritchard and others did find evidence for more sweeps outside Africa.

"A plausible explanation is that humans experienced many novel selective pressures as they spread out of Africa into new habitats and cooler climates," they wrote. "Hence there may simply have been more sustained selective pressures on non-Africans for novel phenotypes."[18] Phenotype refers to the physical trait or organism produced by the DNA, as contrasted with the DNA itself, which is known as the genotype. One obvious example of a novel phenotype needed outside Africa is that of skin color. Africans have retained the default dark skin of the ancestral human population, whereas East Asians and Europeans, descendants of populations who adapted to extreme northern latitudes, have evolved pale skin.

Both within Africa and in the world outside, social structure underwent a radical transition as populations began to grow after the beginning of agriculture some 10,000 years ago. Independently on all three continents, people's social behaviors started to adapt to the requirements of living in settled societies that were larger and more complex than those of the hunter-gatherer band. The signature of such social changes may be written in the genome, perhaps in some of the brain genes already known to be under

selection. The MAO-A gene, which influences aggression and antisocial behavior, is one behavioral gene that, as mentioned in the previous chapter, is known to vary between races and ethnic groups, and many more will doubtless come to light.

Hard Sweeps and Soft Sweeps

Textbooks about evolution discuss favorable alleles that sweep through a population and become universal. There are many ancient alleles that have probably become fixed in this way. All humans, at least compared with chimpanzees, carry the same form of the FOXP2 gene, which is a critical contributor to the faculty of speech. A variation called the Duffy null allele has become almost universal among Africans because it was an excellent defense against an ancient form of malaria. A gene called DARC (an acronym for Duffy antigen receptor for chemokines) produces a protein that sits on the surface of red blood cells. Its role is to convey messages from local hormones (chemokines) to the interior of the cell. A species of malarial parasite known as *Plasmodium vivax,* once endemic in parts of Africa, learned how to use the DARC protein to gain entry into red blood cells. A mutated version of the DARC gene, the Duffy null allele, then became widespread because it denies the parasite access to the blood cells in which it feeds and thus provides a highly effective defense. Almost everyone in Africa carries the Duffy null allele of DARC, and almost no one outside does.[19]

Many other mutations have arisen to protect people against current strains of malaria, such as those that cause sickle-cell anemia and the thalassemias. Sickle-cell anemia occurs with high frequency in Africa, and beta-thalassemia is common in the Mediterranean, but neither has attained the universality of the Duffy null allele within a population. Another widespread but fairly exclusive allele is associated with skin color. This is an allele of KITLG (an acronym for KIT ligand gene) which leads to lighter skin. Some 86% of Europeans and East Asians carry the skin lightening allele of KITLG. This allele evolved because of a mutation in the ancestral, skin-darkening

version of KITLG, which is carried by almost all Africans.[20] A skin-lightening allele of another gene, called SLC24A5, has swept almost completely through Europeans.

But the number of such genes, in which one allele has gone to fixation in one race and a different allele in another, is extremely small and in no way sufficient to account for differences between populations. Pritchard found no cases of an allele going to fixation among the Yoruba, a large African tribe in Nigeria. This has led him and other geneticists to conclude that complete sweeps have been much rarer in human evolution than supposed.[21]

But given that all humans have the same set of genes and that there have been almost no full sweeps that push different alleles to dominance in different races, how have races come to differ from one another? The answer that has dawned on geneticists in the past few years is that you don't always need a full sweep to change a trait. Many traits, like skin color or height or intelligence, are controlled by a large number of different genes, each of which has alleles that individually make small contributions to the trait. So if just some of these alleles become a little more common in a population, the trait will be significantly affected. This process is called a soft sweep, to contrast it with a full or hard sweep, in which one allele of a gene displaces all the others in a population.

Pritchard gives the example of height, which is affected by hundreds of genes, because there are so many ways in which height can be increased. Suppose there are 500 such genes and each comes in two forms, with one allele having no effect on height and the other increasing it by 2 millimeters. An individual's height depends on how many of the height-enhancing alleles he inherits. And that number in turn is determined by the frequency of each type of allele, meaning how common it is in the population. So if each of the height-promoting alleles becomes just 10% more common in the population, almost everyone will inherit more of them, and the average person's height will increase by 200 millimeters, or 20 centimeters (8 inches).[22]

This soft sweep process—a small increase in frequency in many alleles—is a much easier way for natural selection to operate than through the hard sweeps—the major jump in frequency of a single allele—that are often assumed to be the main drivers of evolution. The reason is that the hard sweeps depend on a mutation creating a novel allele of great advantage, which happens only very rarely in a population. In a small population, it may take many generations for such a mutation to occur. Soft sweeps, on the other

hand, act on alleles that already exist and simply make some of them more common. Soft sweeps can thus begin whenever they are needed.

So suppose a group of pygmies were to leave their forest habitat and start herding cattle in a hot climate, where it's advantageous to be tall and thin, like the Nuer and Dinka of the Sudan. The pygmies who were slightly taller would produce more children, and the height-promoting alleles of the genes that affect height would immediately begin to become more common in the population. In each generation, an individual would have a slightly greater chance of inheriting the height-promoting alleles, and the population would quite quickly become considerably taller.

Consider, on the other hand, a trait in which there is no existing variation, such as the ability to digest milk in adulthood. For most of human existence and still in most people today, the gene for lactase is switched off shortly after weaning. To keep the gene switched on requires a beneficial mutation in the region of promoter DNA that controls it. But the promoter region is some 6,000 DNA units in length and occupies a minuscule fraction of the 3 billion units of the genome. In a small population, it might take many generations for the right mutation to occur in so small a target.

Thus it seems to have taken around 2,000 years—some 80 generations—after the start of cattle breeding for the right mutation in the lactase promoter region to appear among the people of the Funnel Beaker Culture, cattle herders who occupied northern Europe some 6,000 years ago. Once established, the mutation spread rapidly and is now found at high frequency in northern Europe.

Three mutations, which differ from one another and from the European mutation but have the same effect, arose independently among pastoralist peoples in eastern Africa and have swept through roughly 50% of the population. In each case, evidently, evolution has had to wait until the right mutation occurred, whereupon the allele grew more common because of the great advantage it conferred.

In sum, hard sweeps cannot start until the right mutation occurs, and then they may take many generations to sweep through a population. Soft sweeps, based on standing variation in the many genes that control a single trait, can start immediately. For a species that undergoes a sudden expansion in its range and needs to adapt quickly to a succession of different challenges, the soft sweep is likely to be the dominant mechanism of evolutionary change. This explains why so few hard sweeps are visible in the human genome. Soft

sweeps are presumably far more common, though at present are very hard to detect. The reason is the difficulty of distinguishing between the minor changes in allele frequency caused by genetic drift and the also minor changes brought about by natural selection through a soft sweep.

The Genetic Structure of Race

It is now possible to understand the structure of human variation, at least in broad outline. Different populations don't have different genes—everyone has the same set. Of the traits specific to one race or another, a few are encoded in hard sweep alleles that have gone almost to fixation, such as the Duffy null allele or some of the alleles involved in shaping skin color, but many more are probably encoded in soft sweeps and hence in mere differences in the frequency of the cluster of alleles that shape each trait.

The fact that genes often work in combination to specify a trait explains how there can be so much variation in the human population and yet so few fixed differences between populations.

Given the importance of allele frequencies in shaping specific traits, it's not surprising that they afford a means of identifying an individual's race. Excluding subjects of a different race is an essential procedure in surveys to detect the alleles that contribute to complex diseases like diabetes and cancer. The idea of these surveys, known as genomewide association studies, is to see if people who are particularly prone to disease are also more likely to carry a particular allele. If so, the allele may be associated with the disease. But the statistics can be confounded if the population being surveyed includes people of more than one race. An apparent association may emerge between the disease state and a particular allele even though the association is really due to some patients belonging to another race, one that naturally has a high frequency of the allele in question.

Medical geneticists have therefore developed sets of test alleles that can be used to distinguish one race from another. Some alleles, particularly those with large differences in frequency between races, are more useful than

others. These race-distinguishing DNA sites are known blandly as AIMs, or ancestry informative markers. Using a set of 326 AIMs, researchers achieved a nearly perfect correspondence between the race that subjects said they belonged to and the race to which they were assigned genetically.[23] A set of 128 AIMs suffices to assign people to their continental race of origin, whether European, East Asian, American Indian or African.[24] (The fifth continental race, Australian aborigines, could doubtless be identified just as easily, but political restrictions have so far largely blocked the study of aborigine genetics.)

With greater numbers of markers, more closely related groups can be distinguished, such as the various ethnicities within Europe.

Some biologists insist that AIMs do not prove the existence of race and that they point instead to geographic origin. But geographic origin correlates very well with race, at least on the continental level.

Apart from genetic markers like the Duffy null allele, found almost exclusively in people of African ancestry, most AIMs are alleles that are just somewhat more common in one race than in another. A single AIM that occurs in 45% of East Asians and 65% of Europeans says that the carrier is a little more likely to be European, but is hardly definitive. When the results from a string of AIMs are combined, however, an answer with high statistical probability is obtained. This is the same general method used in DNA fingerprinting, except that the 14 sites at which the genome is sampled in forensic DNA analysis are not SNPs but variable runs of DNA repeats.

The approach of comparing allele frequencies can even be used with people of mixed race to assign component parts of an individual's genome to their parent's racial origin. When people of different races marry, their children are perfect blends of their parents' genes. But at the genetic level, the chunks of DNA that came from the mother's and the father's races remain separate and distinguishable for many generations. Researchers can track along the chromosomes of African Americans, assigning each stretch of DNA to either African or European ancestors. In one recent study, researchers analyzed the genomes of almost 2,000 African Americans and found that 22% of their DNA came from European ancestors and the rest from Africans, a conclusion in line with several previous reports.[25]

The same study found evidence that African Americans may already have begun adapting genetically to the American environment in the several generations since their ancestors arrived in the United States. The malaria-

protecting genetic variants common in Africans, such as the variation that causes sickle-cell anemia, are no longer a necessity of survival in the United States, so the pressure of natural selection to retain these variants would be relaxed. The authors found some evidence that these variants have indeed declined in frequency in African Americans, while genes that provide protection against influenza have grown more common. The finding, if confirmed, would be a striking instance of evolutionary change within the past few hundred years. A larger and more recent survey, however, has found no evidence of natural selection at work on the African American population since its formation. [26]

Over the last 50,000 years, modern humans have been subjected to enormous evolutionary pressures, in part from the consequences of their own social culture. They explored new ranges and climates and developed new social structures. Fast adaptation, particularly to new social structures, was required as each population strove to exploit its own ecological niche and to avoid conquest by its neighbors. The genetic mechanism that made possible this rapid evolutionary change was the soft sweep, the reshaping of existing traits by quick minor adjustments in the sets of alleles that controlled them.

But what began as a single experiment with the ancestral human population became a set of parallel experiments once the ancestral population had spread throughout the world. These independent evolutionary paths led inevitably to the different human populations or races that inhabit each continent.

Arguments Against the Existence of Race

Readers who are by now persuaded that recent human evolution has resulted in the existence of races may wish to proceed to the next chapter. But for those who remain perplexed that so many social scientists and others should argue race does not exist, here is an analysis of some of their contentions.

Start with Jared Diamond, the geographer and author of *Guns, Germs, and Steel*, who was quoted in chapter 4 as comparing the idea of race with the

belief that the Earth is flat. His principal argument for the nonexistence of race is that there are many different “equally valid procedures” for defining human races, but since all are incompatible, all are equally absurd. One such procedure, Diamond proposes, would be to put Italians, Greeks and Nigerians in one race, and Swedes and Xhosas (a southern African tribe) in another.

His rationale is that members of the first group carry genes that confer resistance to malaria and those of the second do not. This is just as good a criterion as skin color, the usual way of classifying races, Diamond says, but since the two methods lead to contradictory results, all racial classification of humans is impossible.

The first flaw in the argument is the implied premise that people are conventionally assigned to races by the single criterion of skin color. In fact, skin color varies widely within continents. In Europe it runs from light-skinned Swedes to the olive complexion of southern Italians. Skin color is thus an ambiguous marker of race. People belong to a race not by virtue of any single trait but by a cluster of criteria that includes the color of skin and hair, and the shape of eyes, nose and skull. It is not necessary for all these criteria to be present: some East Asians, as noted above, lack the EDAR allele for thick hair, but they are still East Asians.

The single criterion that Diamond proposes as an alternative, genes that confer resistance to malaria, makes no evolutionary sense. Malaria became a significant human disease only very recently, some 6,000 years ago, and each race then independently developed resistance to it. Italians and Greeks resist malaria because of mutations that also cause the blood disease known as thalassemia, whereas Africans resist malaria through a different mutation that causes sickle-cell anemia. The trait of resisting malaria is one that has been acquired secondarily to race, so obviously it is not an appropriate way of classifying the populations. A scholar’s duty is to clarify, but Diamond’s argument seems designed to distract and confuse.

A more serious and influential argument, also designed to banish race from the political and scientific vocabulary, is one first advanced by the population geneticist Richard Lewontin in 1972. Lewontin measured a property of 17 proteins from people of various different races and calculated a measure of variation known as Wright’s fixation index. The index is designed to measure how much of the variation in a population resides in the population as a whole and how much is due to differences between specific subpopulations.

Lewontin's answer came out to 6.3%, meaning that of all the variations in the 17 kinds of protein he had looked at, only 6.3% lay between races, while a further 8.3% lay between ethnic groups within races. These two sources of variation add up to around 15%, leaving the rest as common to the population as a whole. "Of all human variation, 85% is between individual people within a nation or tribe," Lewontin stated. He concluded on this basis that "human races and individuals are remarkably similar to each other, with the largest part by far of human variation being accounted for by the differences between individuals."

He went on to say that "Human racial classification is of no social value and is positively destructive of social and human relations. Since such racial classification is now seen to be of virtually no genetic or taxonomic significance either, no justification can be offered for its continuance." [27]

Lewontin's thesis immediately became the central genetic plank of those who believe that denying the existence of race is an effective way to combat racism. It is prominently cited in *Man's Most Dangerous Myth: The Fallacy of Race*, an influential book written by the anthropologist Ashley Montagu with the aim of eliminating race from the political and scientific vocabulary. Lewontin's statement is quoted at the beginning of the American Anthropological Association's statement on race and is a founding principle of the assertion by sociologists that race is a social construct, not a biological one.

But despite all the weight that continues to be placed on it, Lewontin's statement is incorrect. It's not the basic finding that is wrong. Many other studies have confirmed that roughly 85% of human variation is among individuals and 15% between populations. This is just what would be expected, given that each race has inherited its genetic patrimony from the same ancestral population that existed in the comparatively recent past.

What is in error is Lewontin's assertion that the amount of variation between populations is so small as to be negligible. In fact it's quite significant. Sewall Wright, an eminent population geneticist, said that a fixation index of 5% to 15% indicates "moderate genetic differentiation" and that even with an index of 5% or less, "differentiation is by no means negligible." [28] If differences of 10 to 15% were seen in any other than the human species they would be called subspecies, in Wright's view.[29]

Why should Wright's judgment that a fixation index of 15% between races is significant be preferred over Lewontin's assertion that it is

negligible? Three reasons: (1) Wright was one of the three founders of population genetics, the relevant discipline; (2) Wright invented the fixation index, which is named after him; (3) Wright, unlike Lewontin, had no political stake in the issue.

Lewontin's argument has other problems, including a subtle error of statistical reasoning named Lewontin's fallacy.[30] The fallacy is to assume that the genetic differences between populations are uncorrelated with one another; if they are correlated, they become much more significant. As the geneticist A.W.F. Edwards wrote, "Most of the information that distinguishes populations is hidden in the correlation structure of the data." The 15% genetic difference between races, in other words, is not random noise but contains information about how individuals are more closely related to members of the same race than those of other races. This information is brought to light by the cluster analyses, described earlier in this chapter, which group people into populations that correspond at the highest level to the major races.

Despite the misleading political twist on Lewontin's argument, it became the centerpiece of the view that racial differences were too slight to be worth scientific attention. The assertion left the ugly implication that anyone who thought otherwise must be some kind of a racist. The subject of human race soon became too daunting for all but the most courageous and academically secure of researchers to touch.

A frequent assertion of those who seek to airbrush race out of human variation is that no distinct boundaries can be drawn between one race and another, leaving the implication that races cannot exist. "Humanity cannot be classified into discrete geographic categories with absolute boundaries," proclaims the American Association of Physical Anthropologists in its statement on race.[31] True, races are not discrete entities and have no absolute boundaries, as already discussed, but that doesn't mean they don't exist. The classification of humans into five continental based races is perfectly reasonable and is supported by genome clustering studies. In addition, classification into the three major races of African, East Asian and European is supported by the physical anthropology of human skull types and dentition.

A variation on the no distinct boundary argument is the objection that the features deemed distinctive of a particular race, like dark skin or hair type, are often inherited independently and appear in various combinations. "These facts render any attempt to establish lines of division among biological

populations both arbitrary and subjective," states the American Anthropological Association's statement on race.[32] But as already noted, races are identified by clusters of traits, and to belong to a certain race, it's not necessary to possess all of the identifying traits. To take a practical example of what the anthropologists are talking about, most East Asians have the sinodont form of dentition, but not all do. Most have the EDAR-V370A allele of the EDAR gene, but not all do. Most have the dry earwax allele of the ABCC11 gene, but not all do. Nonetheless, East Asian is a perfectly valid racial category, and most people in East Asia can be assigned to it.

Even when it is not immediately obvious what race a person belongs to from bodily appearance, as may often be the case with people of mixed-race ancestry, race can nonetheless be distinguished at the genomic level. With the help of ancestry informative markers, as noted above, an individual can be assigned with high confidence to the appropriate continent of origin. If of admixed race, like many African Americans, each block of the genome can be assigned to forebears of African or European ancestry. At least at the level of continental populations, races can be distinguished genetically, and this is sufficient to establish that they exist.