

Message in a Genome

As scientists tease out the human genome's secrets, it's easy to seize on our genetic differences, which are small and often inconclusive. But the surprising ancestral connections that our DNA reveals are the big story in the post-genome world.

BY MATTHEW STREMLAU

A CENTURY AND A HALF AGO, AN AUSTRIAN priest conducted an elegant set of experiments that eventually led a grudging world into the first genetics revolution. With a garden of nearly 30,000 pea plants, and meticulous persistence, Gregor Mendel developed the modern concept of the gene. His idea was simple: Observable plant traits, such as stem size or seed coat color, were passed from generation to generation in heritable units called genes. A hundred years passed before it was discovered that individual genes were the instructions for manufacturing proteins. Proteins, along with other molecules in the cell, produce the traits we see in living things every day. In humans, they're responsible for that bald spot we wish we didn't have and the artistic ability we wish we did.

Today, we are in the midst of another revolution in genetics. Around the clock, in laboratories from Boston to Beijing, an army of scientists are decoding any DNA they can get their hands on—DNA from human beings, bacteria, even woolly mammoths whose remains were preserved in frozen ground. In the case of humans, the sequencing of the genome determines the exact order for a particular indi-

vidual of the three billion chemical building blocks that make up the DNA of our 23 chromosome pairs. Superfast DNA sequencers are revealing a vast genetic landscape every bit as exotic as the moon must have looked to the astronauts during the first lunar landing.

While the sequence of a single genome—the complete set of genetic material contained in the chromosomes in each of our cells—shows how human genes are organized, it's the ability to rapidly sequence and compare segments of genomes from thousands of people that is revolutionizing genetics. By examining the natural variation of the human genome, we are identifying genes that play a role in many common diseases. With this knowledge, we may be able to design drugs and personalize medicines for individual patients. And we can use the beacons of genomic variation as a sort of GPS to trace all humans back to a set of common ancestors, most likely in eastern Africa.

Perhaps most remarkably, we can now begin to ask, What makes you you, and me me? We can isolate the variations in our genomes that—along with environmental factors—shape our unique personalities and behaviors. Why am I afraid of heights? Why does my sister hate lima beans while I can't get enough of them? Discovering the answers will be thrilling. But

MATTHEW STREMLAU is a member of the Secretary of State's Policy Planning Staff at the U.S. State Department. He received his Ph.D. in biochemistry from Harvard University.



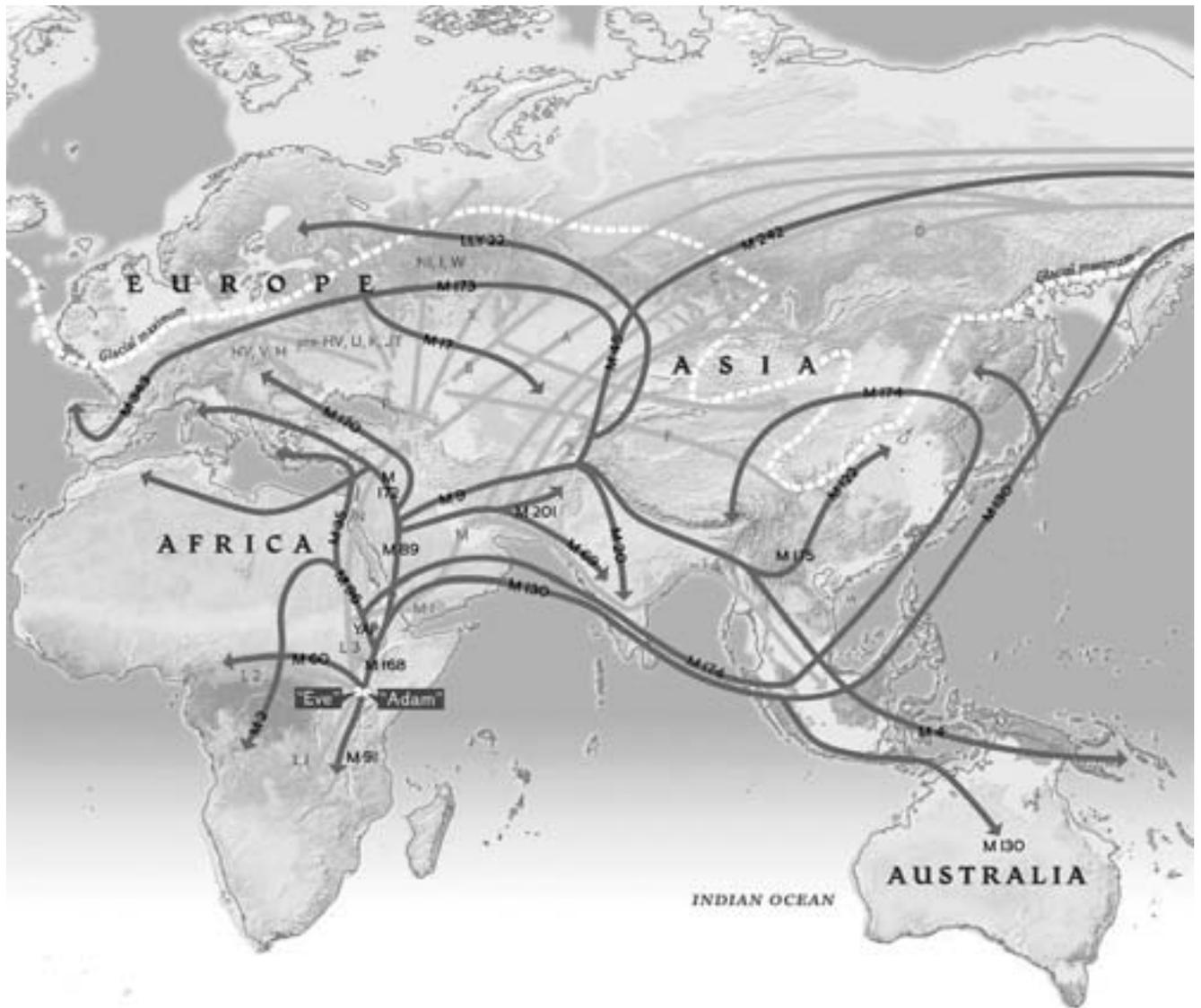
We're not so different, you and me: Despite humans' superficial differences, 99.5 percent of our DNA is the same. Research on human genomes is revealing intimate and sometimes surprising ancestral connections that may change the way we conceive of ourselves.

these discoveries will not come without perils.

Probing our differences, even innocuous ones such as variations in the sense of smell, may resurrect old ideas about genetic determinism—the false belief that our traits and behaviors are determined solely by our genes. This idea gained currency in the latter half of the 19th century, after Charles Darwin's publication of *Origin of Species* in 1859. Sir Francis Galton, Darwin's cousin, argued in his 1865 essay "Hereditary Talent and Character" that each individual's abilities are determined by genetic inheritance, helping to lay the groundwork for the eugenics movement

and its ill-fated efforts to breed better humans.

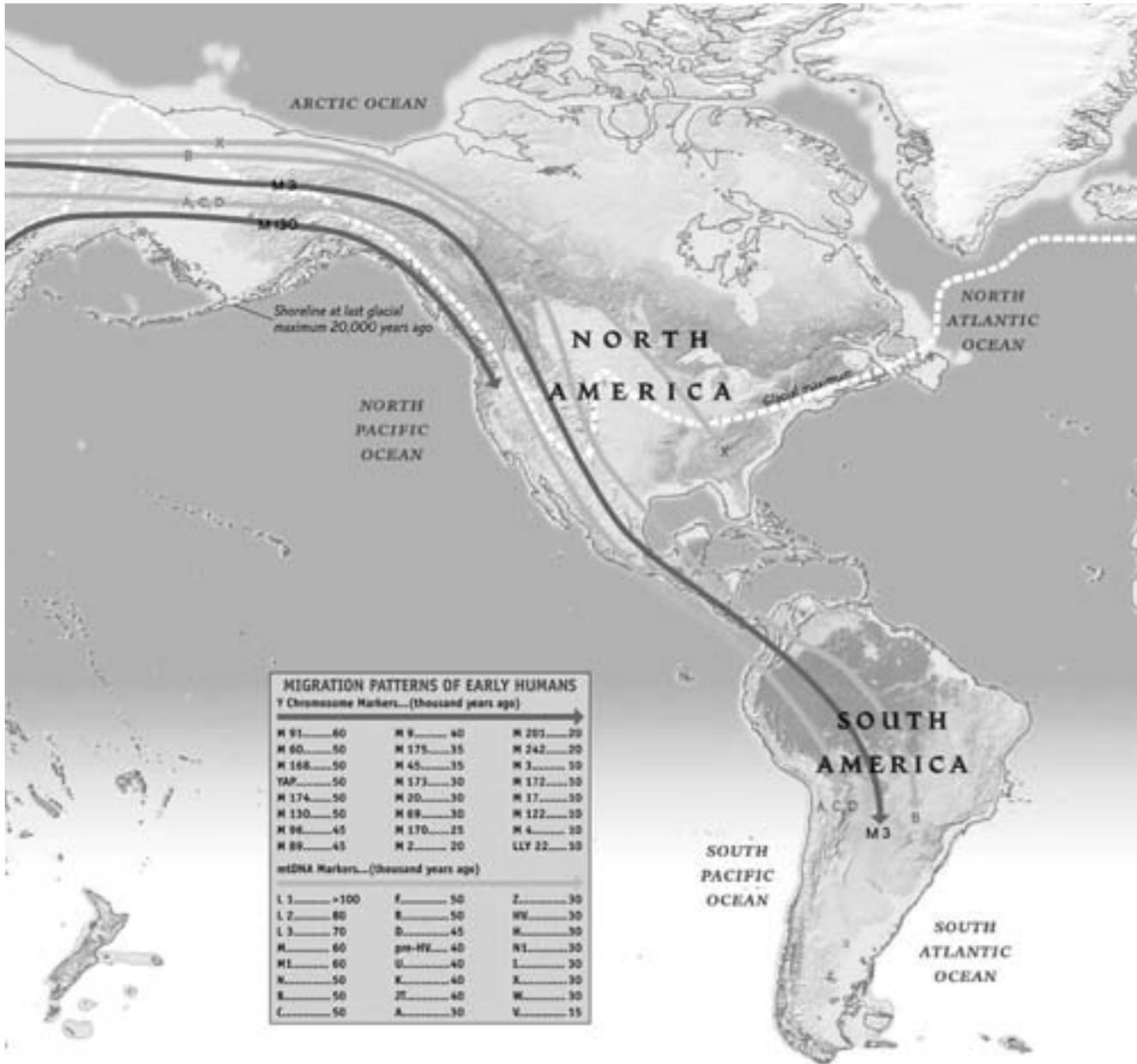
The eugenics movement did not survive World War II, but genetic determinism still has enthusiasts. A few advocates for the death penalty have focused on the so-called warrior gene, a variation of the gene important in regulating the levels of neurotransmitters such as serotonin and dopamine, to argue that some criminals are predisposed to violence and thus incapable of rehabilitation. Others have sought to capitalize on genetic determinism to reduce their own marginalization. In the mid-1990s, activist scientists searched for a "gay gene" that might win greater



social acceptance for homosexuality. But the notion that there is an all-controlling gay gene or warrior gene is flawed. While genes alone are responsible for a few rare disorders such as cystic fibrosis and Huntington’s disease, they are not the last word on most human traits.

Yet we seem unable or unwilling to relinquish the old paradigms of Mendel’s era. As we enter the post-genome world—a world in which we will have easy access to the

information contained in our and others’ genomes—we still tend to believe that variation in one or a few genes is responsible for traits such as kindness or a propensity to develop diabetes. But as we have known for many years, environment plays a major role in our development, and even traits thought to be primarily genetic aren’t likely to correspond to a single gene. In just the past few years, studies have shown how dizzyingly complex this business of the



genome is, suggesting that common diseases are caused by vast networks of genes that interact with one another and with the environment in ways we don't fully understand.

In 2001, the publicly funded Human Genome Project, along with Craig Venter and his colleagues at the private-sector company Celera

Population geneticists are using mutations in DNA to trace the migration of the species since humans first left sub-Saharan Africa some 50,000 years ago. By studying DNA markers in Y chromosomes (passed down paternally) and mitochondrial DNA (passed down maternally), they have plotted who went where and when.

Genomics, announced that they had sequenced a “rough draft” of an entire human genome. (The Human Genome Project published the first com-

plete sequence of the human genome in 2003.) These genomes were composites of DNA from numerous donors and were meant for use only as reference maps. They told us a lot about how the genome is organized, but nothing about human variation. To study that, you need to compare information from the genome sequences of many individuals.

Six years after the first human genome was sequenced, only a handful of individual genomes have been fully sequenced, including those of two Caucasian males, a Chinese male, a Nigerian male, and a Korean male. This past winter, a draft of a genome sequence from a Neanderthal male was announced. There are more genomes on the way (including, finally, a woman's). An ambitious international consortium known as the 1,000 Genomes Project plans to sequence the genomes of 1,200 people from around the world. While the Human Genome Project spent more than a decade and nearly \$3 billion to sequence the first complete human genome, several years from now anyone will be able to receive a genome sequence within a week for less than the price of a good used car.

The human genome contains around 25,000 genes, but that is not all it consists of. In fact, 98 percent of the genome does not contain any genes. This part is often referred to as “junk” DNA. Of course, it's not junk. We simply don't yet understand its function. Much of our junk DNA consists of repetitive sequences and bits of viruses, legacies of infections that our ancestors survived. The junk DNA probably plays an important role in turning genes on and off and also serves as a structural scaffold. Genes are interspersed throughout the genome—sort of like towns along a highway. Some genes are in clusters. Others are separated by great spans of junk.

Over the past few years, scientists have been systematically cataloging all of the common points in the genome that can differ from one person or population to the next. These points are called single nucleotide polymorphisms, or SNPs (pronounced “snips”) for short. The building blocks of DNA, called nucleotides or bases, are denoted by the letters A, T, C, and G. A SNP occurs when a base pair—which

runs together with other pairs to form a sequence—differs between one person and the next.

One SNP familiar to scientists, for example, is associated with Alzheimer's disease. All of us have a gene ApoE that produces a protein whose function is to transport cholesterol and other fats in the bloodstream. Some of us have a SNP in this gene that produces a protein that slightly differs from the “normal” protein. This altered protein does not cause Alzheimer's, but it significantly increases the likelihood that a person will develop the disease.

SNPs can occur anywhere in the genome, both in our genes and in the so-called junk DNA. Geneticists often focus on the SNPs found within genes because it's easier to understand and test their functional significance. But because genes occupy such a small part of the genome “highway,” most SNPs are found outside of genes. These SNPs may play important roles in regulating genes, but the function of most of them remains unclear.

Much of the sequencing work done today is SNP analysis, which is less costly and time consuming than sequencing an entire genome. But many scientists, including Venter, believe that focusing on SNPs is misguided since we don't yet know which parts of the genome are medically relevant. A test that finds a few SNPs that could indicate an increased risk for developing a particular disease may not uncover other genes that reduce the risk. Furthermore, SNPs are just one type of sequence variation in the genome. Certain individuals have more copies of a gene—for example, people with high-starch diets, such as the Japanese, have several extra copies of a gene that helps them digest starch. Other types of variation include insertions, deletions, and inversions of genome sequences.

Compare any two human genomes and you will find that they are 99.5 percent similar. In other words, only 0.5 percent of the genome varies from person to person. You might expect that the farther away a person lives from you, the more different that person's genome is from yours. But the vast majority of genetic variation among people—around 90 percent—is found within continental populations around the world (e.g., Africa, Europe, and South America). That means if you pick a spot on the

genome where your DNA could potentially differ from someone else's, 90 percent of the time it's as likely to differ from your next door neighbor's as from someone living on the other side of the globe.

Only about 10 percent of genetic variation separates population groups. About half occurs between groups within a larger, commonly defined racial group—for example, between Koreans and Japanese, both of whom fall in the broad category of "Asian." The other half occurs between what we think of as typical geographic races such as Africans and Caucasians. Most variation doesn't predominate in any particular group, and genetic boundaries between groups are generally indistinct.

Until around 50,000 years ago, humans evolved in Africa more or less as a group, and variation spread evenly through the population. Today, there is more diversity within Africans as a group than within populations whose ancestors migrated from the continent millennia ago. If you compare two random Italian genomes, for example, they will show less variability than two Kenyan genomes. On the whole, humans are much less genetically diverse than one of our closest primate relatives—the chimpanzees. Compare any two chimpanzees from the same troop, and you'll discover far greater genetic variation than between any two humans. Chimpanzees have been evolving as a species much longer than humans.

The genomics revolution allows us to harness this variation among people to help identify genes that are involved in common diseases or that shape behaviors. A common method for identifying these factors is the genome-wide association study. More than 100 such studies have been conducted so far. The basic idea is to take a group of healthy individuals and compare them to patients suffering from a particular disease. By examining a set of commonly occurring genetic variants, scientists attempt to iden-

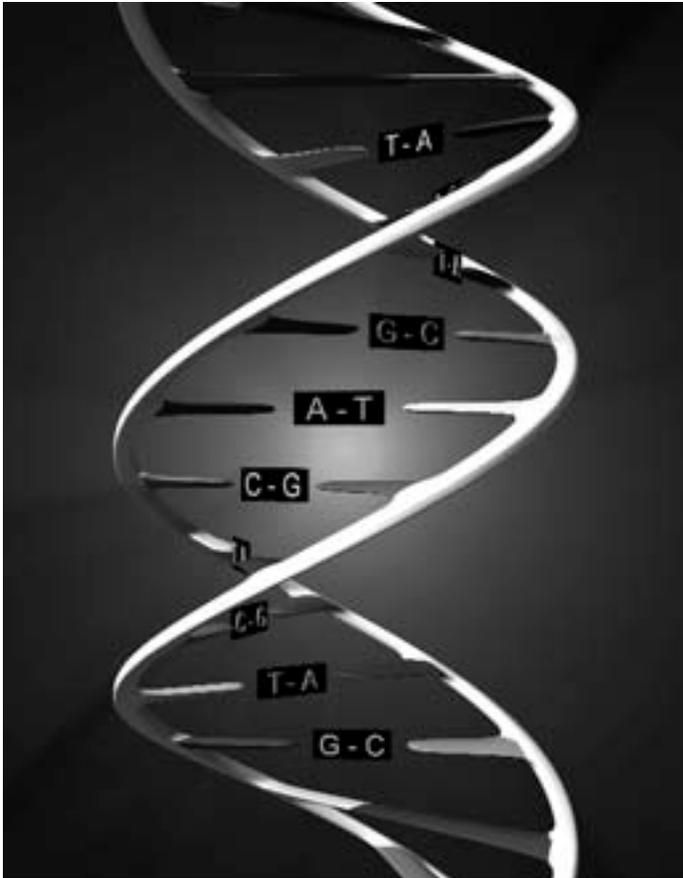
tify ones that appear more frequently in either the disease group or the healthy group. Similar studies can be used to identify the genetic basis of different behaviors. You could, for example, take a group of men who persistently cheat on their wives and compare them to a group who do not. Identify variants in either group, map them to a location on the

ON THE WHOLE, HUMANS are much less genetically diverse than one of our closest primate relatives—the chimpanzees.

genome, and reveal genes that may play a role in determining faithfulness or infidelity.

Ideally, for identification purposes a single mutation in a single gene is the cause of the disease. But this scenario plays out for only a very small number of rare diseases. Most studies of common diseases uncover a large number of relevant genes, each of which makes only minor contributions to the overall disease. Take diabetes, for example. Researchers have so far identified 19 genetic variants associated with type II, or adult-onset, diabetes, and they estimate that they may eventually find anywhere from 100 to 800. The genetic determinants of many common diseases are more numerous and complex than was initially thought. For other diseases, a few extremely rare and difficult-to-detect mutations are the culprits. This past spring, *The New England Journal of Medicine* published a series of articles by well-known geneticists who expressed frustration and disappointment at their inability to hit upon a one-size-fits-all genetic explanation for common diseases, which had seemed within their grasp only a few years ago.

If understanding the role of genetic variation in common diseases is not easy, imagine trying to carry out the same kind of genomic analysis with a complex and difficult behavior that is not easy to define,



The human genome has approximately three billion DNA base pairs, or “letters.” Genome sequencers look for places where the letters vary from person to person.

such as compassion. Unlike diabetes, compassion is difficult to “diagnose” precisely. Once you have defined compassion in some manner and have scanned the genomes of people you believe are compassionate, you will likely identify many variants spread across the entire genome. The vast majority will be rare and produce very subtle effects. And most likely the majority will fall somewhere in the junk DNA, making it even more difficult to understand their function. As you compare the genomes of those presumed to be compassionate to those of the control group, only a tiny fraction of these variants will appear more frequently in one group or the other. Most of the variants will appear with equal frequency in both groups.

Given our current state of knowledge, it will be nearly impossible to pin down the effects and func-

tions of the few variants that may be significant because they are specific to one group or the other. All that we would really be able to say is that in some people, whom we’ve identified as being compassionate using a very vague (and controversial) metric for compassion, we’ve found a genetic variant in the genome that appears with slightly greater frequency (a whopping 11 percent, say) in the group of people we think are compassionate. But we can’t isolate and test the function of that gene because we don’t yet have a method for testing compassion in the lab. Still, don’t be surprised when you open the newspaper and find a headline that announces, “Gene for Compassion Identified in French but not Americans.” Most likely, an overzealous (French) reporter will have homed in on one rare variant and co-opted it to make a sweeping generalization.

There are, of course, examples of one or a few genes that produce dramatic group-specific phenotypes such as skin color or lactose tolerance (a trait common among northern Europeans but much less common among southern Europeans). Recently, a variant of a gene that controls responses to nerve

signals was discovered to be more prevalent in African Americans than in Caucasians. This finding helped solve a long-standing puzzle: African Americans and Caucasians respond differently to beta-blockers, a class of drugs used to treat heart disease and hypertension.

But most human traits, and certainly most complex behaviors, are not the product of one gene but of many. Interactions among many genes make the study of individual differences more difficult. For most traits, extreme and easily observable differences between groups likely would require variation among many genes. Since human migration out of Africa was relatively recent, there hasn’t been enough time for groups to acquire the tens or hundreds of variants that, when combined, produce the types of “diagnostic” differences that would distinguish them from other groups.

Although differences between groups account for only a fraction of the total variability among humans, it is nevertheless possible to accurately classify humans into populations based solely on information contained in the genome. Just last year, scientists showed that by looking at a “bar code” of 500,000 variants in the genome, they could determine with surprising accuracy the geographic origin of people in Europe. Using only this DNA bar code as a guide, they were able to place 50 percent of Europeans within 192 miles of their homes and 90 percent within 434 miles. The variants used for this type of analysis are not functional and do not affect the appearance or behavior of people in the group. But these studies do illustrate that if enough carefully selected variants are considered—and many are needed—group differences can emerge.

In the future, researchers will discover differences in complex traits, such as intelligence, among groups. But because intelligence is shaped by a vast network of genes, variations among different groups will necessarily be subtle and exist along a continuum. If the biological component of intelligence were controlled by a single gene, then the story might be different. But most genetic variation in intelligence—like human variation in general—will be within a particular group, not between groups.

Perhaps the greatest achievement of the post-genome world is the ability to trace genetic ancestry. In 1987, using DNA sequences from the mitochondria—structures within the cell that supply energy—New Zealand biologist Allan Wilson and his graduate students traced the human lineage back to a common female ancestor who lived in Africa 170,000 years ago. This mother of all humans came to be known as “Mitochondrial Eve.” More recently, by analyzing variation in the male Y chromosome, researchers were able

to trace human ancestry back to a roughly common male ancestor. This person is known as “Y-Chromosome Adam.” That’s not to say that Mitochondrial Eve and Y-Chromosome Adam were the only humans living during their time, or that they were the first couple (they lived more than 50,000 years apart). In the case of Mitochondrial Eve, she is the ancestor in whom our

USING ONLY THE VARIANTS in the genome, scientists can determine with surprising accuracy the geographic origin of people in Europe.

female lineages eventually converge, so that we can say that everyone living today is derived from that common mother. Likewise, Y-Chromosome Adam is where our male lineages converge.

As humans began leaving Africa around 50,000 years ago, they spread out. One group journeyed into Europe, another into Asia, and so on. As they entered new environments, they evolved. They also divided themselves up into different racial, religious, and geographic groups. As these social constructions became more pronounced, they created new barriers to gene mixing.

Race and ancestry are often confused. Individual ancestry is a far more valuable predictor of human traits than characteristics such as race. For example, variation in the gene for hemoglobin that can produce sickle cell anemia is common among West Africans, and people usually associate sickle cell disease with being black. But the same variant that produces sickle cell anemia also protects individuals from malaria, so it appears among various groups whose ancestors lived in areas where malaria was common, including Middle Eastern and Indian populations. Many blacks whose ancestors didn’t live in malaria-ridden areas do not carry the sickle cell variant. A person might call himself white, but if he has one great-grandparent who carried a sickle cell mutation, he has a one in eight chance of inheriting that mutation. Thus, a person’s ancestral history—whether or not his or her fore-

bears lived in an area of high malaria prevalence—is a better predictor than race of whether or not that individual will have the hemoglobin variant.

As scientists mine the information of our genomes and develop better tools for tracing human ancestry, our notions about the constitution and boundaries of different groups will become increasingly blurred—and conceptions about the size and number of those groups will change as well. The genomic revolution is highlighting the limitations of our current classification systems. For example, Henry Louis Gates Jr., the well-known Harvard cultural critic, made an unexpected

stopped by to talk about his research. He had recently published a scientific article that was picked up by the popular press—an experience most scientists never have in their entire careers. Lahn and his colleagues had shown that two genes that correlate with brain size were undergoing rapid positive selection. They found certain variants that were more common in Eurasians than in Africans, and hypothesized that these variations in genes important for brain development were related to the migration out of Africa millennia ago. No one disputes the science. However, many members of the

scientific community have criticized the political implications embedded in Lahn's report—that these differences contribute to racial differences in brain size and perhaps IQ. *The Wall Street Journal* ran a front-page article that

INSTEAD OF TRYING to regulate genomic research, we should be working harder to open it up.

discovery when he began researching his genetic history a few years ago. Though he identifies as African American, he found that his ancestry can be traced back not only to Nigeria but also to a fifth-century Irish king, Niall of the Nine Hostages. As much as half of his genetic ancestry is European.

These kinds of genetic revelations will become increasingly common in the post-genome world. A 2004 study found that Europeans are more genetically similar to Asians than they are to other Europeans 38 percent of the time. Several studies of African Americans have revealed up to 80 percent European ancestry. Unexpected relationships won't just be revealed among different racial groups. In the 15th and 16th centuries, Christian armies in Spain converted thousands of Jews and Muslims to Christianity. Genomic sequencing of Christians living in Spain and Portugal today revealed that 20 percent of the people on the Iberian Peninsula have Jewish ancestry and 11 percent have Moorish forebears.

interpreted Lahn's research to mean that he had found a gene that makes certain racial groups more intelligent than others.

After the lecture my Chinese colleagues and I returned to the lab, where we debated whether research on genetically based differences in intelligence should be pursued. Most of them felt strongly that it should. They presumed that once we started comparing genomes, we would discover that the Chinese enjoy superior intelligence—and, living in a relatively homogeneous society, they saw little problem with that. While I agreed with them that we need to understand the biological basis for intelligence, we differed on how to pursue that research so that the results won't be misinterpreted.

So how should we deal with the challenging ethical issues of a post-genome world? Instead of trying to regulate genomic research, we should be working harder to open it up. While we must pay close attention to the impact of genetic discoveries, perhaps the best solution is not more oversight by ethics boards but, rather, investment in training scientists from around the world—particularly the developing world. Today, most genomic research is done by sci-

Several years ago, I worked in a stem cell lab in China. One day Bruce Lahn, a Chinese scientist currently at the University of Chicago,

entists working in the United States, Europe, and Japan. Countries such as China, India, and the Persian Gulf states are beginning to take part, but their contributions are still relatively few. The vast majority of the world's populations are not represented in this increasingly global lab. They lack the resources or training.

We must ensure that they have it. Scientists from other cultures will ask questions we don't. And they may come up with answers we don't like—just as, at times, our research produces answers they don't like. Imagine geneticists from Nigeria poking around Birmingham, Alabama, collecting samples for a study on the genetic differences between white southerners and white New Englanders. That scenario is being played out, in reverse, in Africa right now. Only when people from around the world participate in genomics will we get a complete picture of human variation and its role in determining our complex traits.

We know already that we are not inherently equal. We see innate differences among our friends, family, and colleagues every day. For the survival of our species, inherent genetic equality would have been fatal. Our differences are the critical variations that allow our vast human network to adapt to unexpected stresses such as climate change. Someday we will have the ability to visualize this network, perhaps through a genetic version of Facebook. And the new and unexpected relationships that are revealed may well overshadow the differences—now revealed at a genetic level—that we all knew were there anyway.

There will be voices of alarm that genetic information will be co-opted for a 21st-century eugenics agenda. Groups will be discriminated against based on their genetic sequences, easily obtained from a cheek swab or blood sample, the critics will warn. Perhaps. But race, ethnicity, religion, political affiliation, geography, and sexual preference have been, and will continue to be, stronger bases for discrimination than genetic distinctions. While genetic information may provide ammunition for those who want to divide groups, its more profound long-term

effect may be to blur and confuse our notions of group affiliation.

In the future, efforts to sequence complete individual genomes—rather than a subset of common variants within the genome—will be critical for creating a complete collage of human ancestry. With the full picture, we'll probably come to see that everyone is both a genetic winner and a genetic loser. While you may have some highly desirable genetic traits, you will also discover some you wish you didn't. Furthermore, what's beneficial today might not be beneficial tomorrow. For example, northern Europeans have acquired a mutation in a key receptor for HIV that disrupts the ability of the virus to enter the cell. As many as 20 percent of northern Europeans have this mutation. In other groups, such as Africans and Asians, the mutation is extremely rare. There's no question that, in the midst of today's global HIV epidemic, this mutation is beneficial. But tomorrow it might be detrimental. The mutation, while protecting against HIV, ablates the ability of the receptor to function normally. At some point, humans might find a crucial need for this receptor, and those who have it will benefit.

In today's globalized world, where technology is breaking down geographic barriers, people from different groups are mixing with one another more than perhaps at any time since humans first migrated out of Africa. As variants in our genes shuffle around the globe, the genetic differences between groups will diminish. Perhaps, one day, all humans will again be a single race—as they existed many thousands of years ago, before they left Africa.

The world is really a vast genetic network of 6.7 billion people. As more genetic sequences of individuals become available, we'll discover new relationships and relatedness. Many of these genetic relationships will be unexpected. Our definition of race may change. Questionnaires that sequester race into the simplistic categories of Caucasian, Asian, Hispanic, and African will feel confining and outdated. While some undoubtedly will try to discriminate against groups based on genetic information, they will not be able to erase the overwhelming amount of shared ancestral history that the post-genome world will allow us to see. ■