This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

Perspectives on Human Variation through the Lens of Diversity and Race

Aravinda Chakravarti

McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Correspondence: aravinda@jhmi.edu

Human populations, however defined, differ in the distribution and frequency of traits they display and diseases to which individuals are susceptible. These need to be understood with respect to three recent advances. First, these differences are multicausal and a result of not only genetic but also epigenetic and environmental factors. Second, the actions of genes, although crucial, turn out to be quite dynamic and modifiable, which contrasts with the classical view that they are inflexible machines. Third, the diverse human populations across the globe have spent too little time apart from our common origin 50,000 years ago to have developed many individually adapted traits. Human trait and disease differences by continental ancestry are thus as much the result of nongenetic as genetic forces.

Half-jokingly, Gwen Ifill, the noted American journalist and newscaster, told the Smithsonian audience, "In no universe is President Obama white!" (Smithsonian's National Museum of Natural History 2013). Her comment came in response to my genetic argument that given the President's white American mother and Kenyan father, he could just as well be called "white" as "black."

This friendly exchange exposed the essential conundrum surrounding the contemporary meaning of labels, classification, and the notion of race. We humans have, since time immemorial, sorted and classified each other into numerous categories based on language, culture, and appearance (en.wikipedia.org/wiki/Race_(human_classification); Blumenbach 1775). Irrespective of how these groupings were decided on or justified, such classification has been a cultural exercise: the basis for self-identification through the identification of others. Genetics, being a modern science, has come to this scene much later. Genetics has much to say about the recent and remote history of our species and our individual ancestries, as well as the potential to support or refute our existing classifications. What genetics says about our history cannot be wished away. At the same time, our cultures have a strong voice in how we view ourselves and view others. This cultural view cannot be wished away either. President Obama's self-identification as "black" is not based on his personal gene accounting but rather a nonchoice given American social convention (the "one-drop" rule) and his personal history. That was Ifill's point. What we, and others, call us depends on both our genes

Copyright © 2014 Cold Spring Harbor Laboratory Press; all rights reserved
Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a023358

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

and our society. I am a Bengali-American, the duality and hyphen being equally important to my identity over and above my genes.

So, what are we to think of the existence of human "races" in this "genomic age" and why is there still so much controversy? (Koenig et al. 2008). By any account, we cannot ignore the idea of human races. It is in daily common use, a basis of self-identification and for many a key to their social identity. Without quibbling about word usage and specific meaning, race is also the basis for governmental statistical accounting and political action. Although the word race finds daily use in the United States, and, increasingly in Europe, similar controversies surround other classifications of humans, such as caste or tribe, despite the diverse origins of each term (Thapar 2014). The most controversial aspect of such classifications, in my opinion, is not whether they are biological or not, but rather the imputation of wholesale traits[1] and attributes to these groups so defined. Almost without exception, the characteristics displayed by one's own group are deemed positive and implicitly valued, whereas those of other groups are deemed negative and are undervalued. The group defining and possessing the valued features is also invariably that group that is culturally dominant and politically powerful.

Genetics has always been an ideal fuel for this fodder. It is a science that examines the biological basis for trait differences, and has recently allowed us to make tremendous strides in our understanding of human disease (e.g., why some individuals have muscular dystrophy and others do not) and differences (e.g., why some individuals can digest milk and others cannot). The argument goes that if groups can differ in traits such as lactose intolerance (most Europeans and some Africans are tolerant, whereas others of the world are not) and malarial susceptibility (many Africans, Middle Easterners, South Asians, and East Asians have some protection, whereas the rest of the world does not) owing to specific gene differences, they, in all probability, also genetically differ in a host of health-related traits. But, why should this principle be restricted to health-related features? Some have argued, why not genetic differences in intellectual ability, industriousness, the facility for democratic institutions, or aggressive behavior (Wade 2014)? In the past 100 years of genetics, many in the field have advertently and inadvertently engaged in considerable speculation as to these last possibilities, the biological underpinnings of *any* difference. But what is the evidence that these metafeatures are genetic? I know of no science that can prove the genetic underpinnings of these broad social differences. In contrast, I know of plenty of evidence that argues against it being the case (Chakravarti 2010). As the physicist Neil deGrasse Tyson recently quipped: "You get to say the world is flat because we live in a country that guarantees free speech, but it is not a country that guarantees that anything you say is correct" (deGrasse Tyson 2014).

Over the next several pages, I would like to tell you what we know of human diversity and how we came to be who we are today: in short, our singular genetic heritage and history (Fig. 1). I will outline what we know today about genetic differences across contemporary human populations and how these differences can sometimes account for the human trait differences we observe. I will not opine on whether race exists or not; it does so in a very real sense. Instead, I will discuss what modern genetics says or does not say about the genetic meaning of race. Such work, together with increasing knowledge of how genes function at the molecular level, is giving us new insight into how genes influence our traits. But we remain vastly ignorant; so wild speculations about the genetic nature of many human attributes reside beyond the realm of today's science. Importantly, the notion of the gene, in the minds of most—many geneticists included—is one of an inflexible machine with deterministic outcomes. As the science advances, we increasingly find that the effects of genes are highly modifiable, dynamic, and subject to external influences (Chakravarti 2010). Indeed, why only genes? Even their ultimate products, structures such as our brain, are highly modifiable and dynamic (Kays et al. 2012).

[1]The word trait is used to indicate a distinguishing quality or characteristic; in its genetic flavor it means the "phenotype" or the manifesting feature of our genes ("genotype").

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

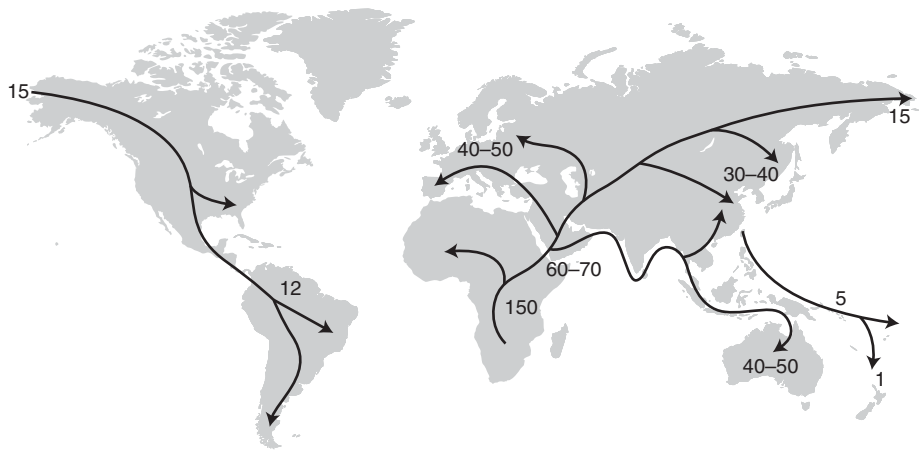

Figure 1. The long trek of our ancestors from the beginnings in Africa ~150,000 years ago to their emergence out of Africa about 50,000 years ago to colonize the Levant, Europe, Asia, Australia, Americas, and eventually, Oceania. (From Gluckman et al. 2009 [Fig. 6.6, p. 142]; reprinted, with express permission, from the authors in conjunction with Oxford University Press © 2009.)

Overlaid on top of all of this are the vast demographic changes human populations are increasingly experiencing, driven by increased communication and movement, and the shedding of past cultural divisions. These changes profoundly affect the distributions of genes across humanity and the melding of what were once population attributes. The science of genetics will be crucial to understanding how human history unfolds, and I predict that most of our current prejudices will turn out to have no biological basis.

WHY ARE WE NOT ALL THE SAME?

There would be nothing to argue about if humans were not different from one another. Of course, there are the rare exceptions of identical twins, but even though they have identical genomes they can on occasion show different traits. This perceptible difference within a collection of similar things extends across all of nature. Science is possible only because these differences exist and is driven by our continual quest to find out how and why they arise. Genetics is a young science, its 100-year history arising from the quest to find how and why biological differences arise and how they are maintained. The "how" was first answered by Gregor Mendel's experiments with peas and the "why" by Charles Darwin's wondrous voyage to the Galapagos (Provine 1971). We persist in continuing to answer these questions in ever more detail because our current understanding, despite being solid, is very incomplete. We are acutely aware of what is not true but often on shaky ground about what is.

We differ in traits because of the biological processes that produce us. Some of this is genetically encoded and some of it is environmentally induced or modified. Each of us develops via a genetic program that is encoded by the sequence of A, C, T, and G bases in the DNA that makes up our genomes. Each of us inherits two genomes, one from each of our parents. This genetic program defines each of us uniquely and is more similar between any two members of the same species than between members of two related species. Thus, human genomes are more similar to one another than any one of our genomes to that of our great ape relative the chimpanzee. The differences do not stop there. Each of us is acutely aware of individual-to-individual differences between humans and the greater similarity between any one of us and our family members. Genetics provides a singular explanation for both observations. The closer the relationship between two persons the more similar their

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

genomes and the greater the similarity of their traits. The dissimilarities between our genomes are owing to constantly arising mutations in our DNA. The majority of these are never transmitted to subsequent generations (they are lost), but a minority persists over time and across the generations. These surviving variations are the currency of modern genetics. Two individuals are related by virtue of sharing one or more common ancestors from whom they have inherited a small segment of their genome. The more remote the common ancestor(s), the less the fraction of the genome shared. Thus, we share 1/2 of our genome with each of our parents and siblings, 1/4 with each of our four grandparents, 1/8 with each of our first cousins, and so on (gcbias.org/2013/12/02/how-many-genomic-blocks-do-you-share-with-a-cousin).

The genetic mutations that persist among us, called genetic variants or polymorphisms, can occur at various frequencies within a population; some are rare and unique to a family, whereas others are common and have spread throughout humanity. These genetic variants can be assessed to evaluate how different two individuals are, and therefore figure out their relationship. The first human genetic variation identified was the ABO blood-group system in 1900, which was used almost immediately for assessing close relationships such as paternity. Today, technological advancements allow us to examine the entire genome in exquisite detail and to identify essentially all of such genetic differences. Typically, when one compares two copies of the human genome, say the maternal and paternal copies in any one of us, we find one of every 1000 bases to be different. Because the human genome is three billion bases long, that represents three million differences. Like the Hubble telescope that has allowed us to see deeper into space and time, new genomic technologies can identify all of these differences today and allow us to detect ever more remote relationships, and ancestries to 100,000 years before present or more, and do so on a global scale. The consequent stories of what these similarities and differences mean for similarities and differences in human traits and diseases are only in their infancy.

THE BIOLOGICAL BASIS FOR VARIATION IN HUMAN TRAITS AND DISEASES

Friar Gregor Mendel was the first geneticist (Orel 1996). He was deeply interested in the question of how differences in plant characteristics arise and how they are propagated. His now famous experiments, using simple observable traits (plant height, seed shape, flower color) of the pea plant, allowed him to infer that trait variants were owing to differences in separate "factors," that these factors existed as pairs in individuals, and that one of each was transmitted to each offspring, randomly and independently of other factors. Mendel's factors are today's genes. A fact not appreciated is that Mendel performed many other similar experiments with other traits and other plants and failed to uncover similarly clarifying principles (Orel 1996). This is not to say his rules of inheritance governing genes are incorrect; these apply to all genes. Rather, some genes have overwhelming effects on a trait and so the trait inheritance patterns are simple, the so-called Mendelian traits, such as those of plant height, seed shape, and flower color in the pea plant. Other genes, however, exert their effects in concert with numerous other genes, none of which overwhelm the others and thus have more complex patterns of inheritance. The fact is that Mendelian inheritance of "traits" is very much the exception, not the rule (Chakravarti 2010). The failure to understand this key feature led to considerable and bitter controversy in the early genetics literature (the Mendelian-biometrical debate) when some held that metrical traits such as height were not inherited but rather their variation arose solely from environmental differences (Provine 1971).

Eighty years hence, we are considerably more informed as to the nature of non-Mendelian inheritance. Much of the science of genetics has advanced from the experimental use of Mendelian inheritance to uncover its biological and molecular underpinnings. We also know that although some traits arise from the actions of two or at most a few genes, the vast majority of traits are genetically complex arising from the actions and interactions of numerous, hundreds

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

or even thousands of genes. We know this because contemporary genetics allows us to map the locations of the individual genes contributing to a trait and, for most, a role for hundreds of genes has been revealed (Lango Allen et al. 2010). We also know that many more unmapped genes exist and that the individual-to-individual variation in a trait also involves the contributions from other domains, namely, environmental and epigenetic factors. The term environmental as used in genetics is both broad and nonspecific, and can include everything from lifestyle (diet, exercise) to ecological (weather, altitude) to social (income, education, healthcare access) and cultural (diet, belief systems) factors. The term epigenetic is also broad and includes a whole host of cellular processes that can direct the actions of genes without being dependent on the sequence of a specific genome. There is also increasing evidence that cellular (genetic) outcomes are not deterministic but inherently dynamic and stochastic. Genes do not determine only one outcome, but a range of possible ones. Finally, consider that biological effects cannot be arbitrary but are both canalized (restricted to certain possibilities) and built to preserve homeostasis (physiological regulation maintaining more or less constant internal conditions). There is one more arbiter; evolution decides which of the many genetic changes that occur within our genomes will be retained and which ones will be culled depending on whether the change is beneficial or not.

The variation in any trait, or for that matter disease susceptibility in any species including our own, is the result of many factors (genetic, epigenetic, environmental) each of which can be further divided into multiple subfactors. It is no surprise that traits can be inherited in a complex manner because beyond genes, whose inheritance patterns we understand, the epigenetic and environmental factors can also be "inherited," albeit according to rules still not understood (Cavalli-Sforza and Feldman 1981; Jirtle and Skinner 2007). It is then also unsurprising that Mendelian inheritance of traits is rare, because these represent the unusual singular effects of one gene that overwhelms the extant nongenetic variation.

Geneticists have long been interested in the precise genetic architecture of "complex traits." The first step is assessing a trait's genetic component. The top-down or classical approach has been to compare traits among relatives, because we have long known how much genetic information relatives share (e.g., 50% between siblings) without knowing individual genes. This allowed us to estimate the proportion of variation that is genetic, a proportion called heritability. The concept of heritability has been of great practical utility in plant and animal breeding, as a guide for choosing which strains to develop for improvement of yield and performance-related traits. The heritability of numerous human traits and diseases has also been measured, often repeatedly. Although some traits have high heritability, like height (>80%), the overwhelming majority has low to moderate heritability (30%–50%) (Vinkhuyzen et al. 2013). The specific identification of the genes that explain this heritability, by contemporary bottom-up approaches in which the entire genome is systematically investigated, has been notoriously difficult, however (Lango Allen et al. 2010; Vinkhuyzen et al. 2013). This is sometimes referred to as the "missing heritability" problem.

There are many reasons for this apparent failure. First, our study samples are yet too small and not diverse enough. Second, our technological approaches are insufficient to recognize the vast network of gene interactions that may be principally important. Third, heritability estimates exaggerate the effect of genes because most studies cannot distinguish genetic from social or cultural sharing; family members share much more than genes (social, cultural, and dietary factors). Fourth, heritability is a relative measure of genetic versus nongenetic contributions. Thus, simply increasing or decreasing the environmental part of the variation can alter the apparent role of the genetic part. Well-known examples, such as the height increase seen from improved nutrition in the absence of any genetic change or the dietary treatment of phenylketonuria from birth to prevent intellectual disability, show that the actions of genes can be mitigated by nongenetic interventions.

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

WHY GENES TELL STORIES

Almost every human gene, when its genomic sequence is compared across individuals, shows variation. Each such sequence is a palimpsest, recording all changes from mutations that have survived until today. Some of these changes are unique to particular individuals, perhaps even one, whereas others are present in many of us. Because all humans belong to a single family tree—on average any two of us share 99.9% of our genomes—the fraction of sequence difference between any two genomes indicate how far back in time they had a common ancestor. Different segments of the genome are shared with different common ancestors; so the fraction shared or different between two genomes varies along its length. Consequently, examining the entire genome, as we can do today, is more informative than studying only one bit, such as the maternally inherited mitochondrial genome or the paternally inherited Y chromosome. The latter are informative nevertheless because they allow us to make inferences about our maternal and paternal lineages, respectively. Because all genetic changes accumulate over time, our genomes thus provide a history of how we, as individuals and as a species, came to be. Today we can compare individual genomes to infer our relationships, how far back in time we shared one or more common ancestors, and with increasing precision because of limited mobility of our ancestors, where our forebears were geographically located.

Genomic technologies, genomic sequencing in particular, have opened the door to recovering our individual and collective genetic histories, and, therefore, in concert with other sources of information, to uncovering details of our prehistory. In this sense, genes tell compelling stories about each of us as well as our shared humanity. This is a truly remarkable scientific and social achievement. There are many aspects of these stories that are uncertain and will require revision in the future. Nevertheless, some compelling and surprising truths have emerged. The most important of these is that contemporary humans are a remarkably young species and we all belong to a single family tree that arose from common ancestors a little more than 150,000 years ago (Pääbo 2014). Modern humans came to be in the last few minutes of the last hour of the last day if all 14.5 billion years of cosmic evolution were compressed into one year. If we were the common bacterium *Escherichia coli*, then this would correspond to a mere 3 months of our life. The story of human diversity, why we look as diverse as we seem to, needs to be told with this truth in mind.

WHAT IS RACE?

The Oxford English dictionary defines human races as the "major divisions of humankind, having distinct physical characteristics," and also as a "group of people sharing the same culture, history, language, etc." Biologists have had a more specific definition of race, one not conjured with human diversity in mind. The evolutionary biologist Ernst Mayr wrote that a race is "an aggregate of phenotypically similar populations of a species inhabiting a geographic subdivision of the range of that species and differing taxonomically from other populations of that species" (Mayr 2002). This definition has more to do with biogeography and taxonomy. However, there is an implicit assumption of both transmission and permanence of such taxonomy, and biologists impute the existence of some fundamental genetic and evolutionary difference between groups termed races. If one believes in evolution and modern genetics, and a common tree of life, the conclusion is inescapable that some members of a single species will be more different than others; additionally, close relatives of each of these members will be more similar to their closer rather than their more distant kin. It is unsurprising that this is true for humans and that our many attributes, including physical features, show this pattern. The precise pattern of sharing is a result of our specific evolutionary history and these differences are written in our genes and propagated through them. The construction and existence of human races in this regard, quite apart from social and cultural meanings, would not per se be controversial. It is controversial today because, over the past few centuries, both experts

 Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a023358

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

and nonexperts alike have brought in new and corrupted meaning that is not inherent in the biological concept. Discussions on human race, and caste, are difficult and incendiary today because their subtext is that human genetic differences are not neutral but either advantageous or disadvantageous and, tragically for human history, the corollary is that some groups have mostly advantageous attributes, whereas other groups have largely disadvantageous ones (Herrnstein and Murray 1994; Koenig et al. 2008; Wade 2014). Genetics has provided a second, more pernicious, corollary. Because some of these traits might be "genetic," these differences are transmitted at conception and so are biologically permanent (Herrnstein and Murray 1994; Wade 2014). The implication is that some groups have a genetic advantage. There have never been any empirical data to support these claims, and, moreover, the survival of "diverse" human groups is prima facie evidence of each of our groups' evolutionary success (Fraser 1995).

We humans must have always named and classified each other as we came across our evolutionary kin. However, the rise and expansion of the modern concept of human races had to wait for the conquests by colonial powers that brought Europeans into direct contact with many groups who were different with respect to language, culture, and even physical features. It should be remembered that this was always an asymmetrical and unequal rendezvous, favoring the colonizer and disfavoring the colonized. The notion of human races arose in this background from studies of comparative anatomy of human skulls by the German physician and naturalist Johann Blumenbach in the 18th and 19th centuries (Blumenbach 1775). He classified these skulls, and thereby humanity, into five major classes—Caucasian, Mongolian, Malayan, Ethiopian, and American—and began the horrid practice of providing color aliases (white, yellow, brown, black, and red). He correctly concluded that "individual Africans differ as much, or even more, from other individual Africans as Europeans differ from Europeans." Blumenbach, like his contemporaries, believed in the "degeneration hypothesis," which is that humans were originally "Caucasian" and that other races were the outcomes of environmental degeneration (e.g., through exposure to sunlight). Despite the cultural biases he began with, Blumenbach was far more generous than his contemporaries with his view that Africans were not intellectually lesser than their European counterparts (Blumenbach 1775). Subsequently, despite other investigators classifying humans into anywhere from two to 63 races, no less an authority than Charles Darwin opined that "it is hardly possible to discover clear distinctive characters" between human races, because they "graduate into each other" (Darwin 1871).

The political and economic rise of Europe, and then the United States, in the 19th and 20th centuries, fed many bogus ideas into what evolved as "scientific racism" (Fredrickson 2002). There were parallel developments dealing with caste differences in India, although this is a much older classification. These studies had in common the examination of selected traits and attributes, deemed to be hereditary, in turn justifying the conclusion that there was a well-defined value hierarchy inherent in our species, with some groups much better endowed than others. The new emerging concepts of genetics added a new dimension; heritability assured that both the well- and less-well endowed continued to remain so. These beliefs—and they are largely beliefs of the perpetrators because so much of their data has been subsequently shown to be selectively used, manipulated, and outright fraudulent—led to a long period of eugenics both in the United States and Europe and the subsequent rise of the concept of inherent group superiority. This had disastrous consequences for Jews, Gypsies, intellectuals, nonconformists, and the mentally ill, among others, and biased immigration policies in many countries, including the United States. Unfortunately, geneticists had no small role to play in this crime of historical proportions (Witkowski and Inglis 2008). As a number of the authors in this collection describe, there is still a continuing tendency to conflate all manner of group differences with gene differences (Herrnstein and Murray 1994; Cooper 2013; Duster 2014; Wade 2014).

In the next section, I will turn to a modern accounting of human population variation and

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

what it may say or not say about human races. Irrespective of that answer, one aspect is certain. Modern genetics research does not support the contention that one group or another has all of the positive traits, that we understand the genetics of complex traits sufficiently well enough to know that group differences in traits mean majorly group differences in genes, or that traits with a genetic component have fixed, inviolate, permanent, and unmodifiable effects. Even if we were to defend the idea of human continental ancestry or race it would be impossible to defend the assertion that they are inherently unequal. The remarkable feature of human evolution and adaptation is the widespread commonality of highly advantageous features (speech, cognition, culture) throughout humanity and the less frequent evolutionary innovations that occurred locally (pigmentation).

A BRIEF SYNTHESIS OF RECENT HUMAN EVOLUTION

The evolution of hominids leading to *Homo erectus*, 1.5 to 2.5 million years ago, and then *Homo sapiens* in Africa, is now well established. Although *H. erectus* existed outside Africa, the evidence is very clear that we are all descendants of groups from the African continent. *H. sapiens* first appeared there no earlier than ~300,000 years ago (Klein 1989). The subsequent history, evident in only fragmentary form through fossil remains, is where genetics has been indispensable (Cavalli-Sforza et al. 1994; Pääbo 2014).

The widespread discovery of gene variation in the 1960s immediately prompted studies to assess their relative relationships across human groups. The first study reconstructing human evolution using data from living groups was by Cavalli-Sforza and Edwards in 1964 (Edwards and Cavalli-Sforza 1964). This landmark study produced a "tree" in which extant populations arose through independent evolution by splitting from a common ancestral group that also produced a sister group, and so on. This study yielded two major findings beyond the specific relationships between groups. First, geographic proximity reflected greater genetic similarity across all groups, with the largest difference being between African and Australian Aboriginal samples. This suggested that human colonization occurred through successive and serial migrations. Second, anthropometric measures and skin color showed a very different set of relationships, for example, a close association between African and Australian Aborigines, unrelated to geography but dependent on climate (Cavalli-Sforza and Edwards 1964). Another landmark study by Richard Lewontin in 1972 went on to show that the majority of human genetic variation, on average 85%, existed within any group and that intergroup differences were relatively minor, with the largest being between African and non-African groups (Lewontin 1972). These were not isolated controversial studies but rather the beginning of an onslaught of investigations, using successively larger and larger numbers of genes and humans, which have produced a single, consistent genetic narrative of human history (Cavalli-Sforza et al. 1994).

All early studies of human evolution compared the features and relationships of populations not individuals. In other words, these studies compared the relationships between frequencies of gene variants and not the genomes of individuals. This distinction is critical because past studies depended on the definition of a population. Is it defined by language, culture, geography, physical appearance, caste, or "race?" The definitions, of course, could skew the results one way or another. Of course, populations defined by known specific differences can be different at the genetic level. This is why Allan Wilson's 1987 study of individuals and their mitochondrial genomes is a significant departure from the past (Cann et al. 1987). His research accomplished four major goals. First, they studied individual genomes and not ensemble frequencies; second, they clearly showed that the human evolutionary tree had two major branches, one composed of African mitochondrial genomes only and the other comprising all humans including Africans; third, except for Africans, all other individuals from the same population had multiple origins; and fourth, they dated the common mitochondrial ancestor of all humans to be less than 200,000 years ago.

 Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a023358

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

There have since been many genetic studies using ever-increasing types and numbers of genetic variants and culminating in contemporary studies involving whole genome sequencing of diverse collections of humans. The essential conclusions of the findings described above have now stood the test of time. The single story of human evolution in the last 150,000 years is that all of us today are the descendants of early *H. sapiens* in Africa. A small group of these ancestors migrated "out of Africa" about 50,000 years ago and have since colonized the rest of our globe with each new group serially colonizing new unexplored geography. This is how we came into the Levant (Middle East) and then to both Europe and Asia and its subcontinent, and beyond into New Guinea and Australia. This is also how we went to remote parts of Siberia and came to colonize the New World about 15,000 years ago, which was until then hominid free. Our latest forays, only 2000 years ago, have been into Oceania (Fig. 1). These journeys provide the explanation for the pattern of quantitative differences that Cavalli-Sforza, Lewontin, and Wilson first brought attention to. Yes, there are differences in genetic variation at the continental level and one may refer to them as races. But why are continents the arbiter? If humans have had this single continuous journey disobeying continental residence—and as evidence we have the continuous distribution of genetic variation across the globe, not discrete boundaries like political borders—where do we divide humanity and why? (Weiss and Lambert 2014). All humans, without exception, are one species that has only very recently dispersed, with each population being more related to its proximal geographic neighbors. If we do look, behave, and have features that distinguish us markedly from one another, then these are differences that have arisen and amplified only over the last 50,000 years (2000 generations) and, quantitatively, are very small—only one part in 1000 bases. As a comparison, consider that chimpanzees and humans diverged from a common ancestor more than five million years ago (200,000 generations).

Of course, there are many more details to fill in and we wish to have far greater resolution of the history we already know. In this collection, scholars of genetic variation and evolution in the major geographic regions of the world have outlined both known and more recent genetic studies (Gomez et al. 2014; Majumder and Basu 2014; Ruiz Linares 2014; Veeramah and Novembre 2014). Their research, and the complementary works of others, allows us to make three major novel inferences. First, human populations were not always large and, probably, were almost always small. Second, the current abundance of a group is not a reflection of its past size. Third, human populations are seldom homogeneous and are highly admixed.[2]

We academics and nonacademics alike like to associate the genome only with its biological properties. However, our genomes and genetic variation between them and peoples are also the result of who lives, who dies, and who leaves behind how many offspring (i.e., demography). In other words, our genomes carry the record of both biology and demography, although teasing these apart is neither trivial nor easily corroborated by independent sources of information. In fact, contemporary whole genome genetic variation data emphasize the greater imprint of demography than biology in our genomes. As mentioned above, the chief conclusion of many genetic studies is that only a small group emerged from Africa to colonize the world and each successive colonization involved small numbers of founders as well (Gutenkunst et al. 2009), sometimes in the hundreds. This speaks to the limited amount of variation each founding group carried (which was also a subset of that in its parent group) and the constant threat of extinction to such a small band. Our eventual success is often chalked up to crucial adaptations; however, it is not implausible that there were many attempts, many catastrophes, and we are simply the lucky survivors. Of course, long-term survival came from population expansions, an intrinsic feature of human, and all other, evolutionary success. As other authors

[2]Admixture is used here in the genetic sense, in which extant genes and individuals arise from two different ancestries, as historically occurred with the colonization of the Americas by Europeans.

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

of this collection have documented, human population sizes rose within Africa at the time of our early ancestors, and then outside Africa as their descendants spread across the globe. These population increases must have depended on many environmental and chance factors as well as the inclusion of new arrivals (immigration). These factors are neither stereotypical nor orderly, once again emphasizing the strong effects of chance. A group populous today was not necessarily originally so; on the contrary, they may have been ever so close to extinction. Finally, although in the minds of many genetics is associated with homogeneity, human evolution is nothing else but a story of admixture and heterogeneity. There is now ample evidence in the genetic structures of the peoples of Africa, Asia, Europe, and the Americas that all extant humans are admixed (this collection). The new data are only now revealing whom we were admixed with, whether such admixture was common or rare, and when it occurred. Indeed, it is this story of mixing that is so at odds with the classical view of human group identity (Reich et al. 2009). We tend to think of admixture as a feature of modern times and, as has frequently happened, in terms of subjugation of one group by another as has occurred many times across human history and geography. However, this must have occurred even in the remote past; as human population density increased, there must have been a greater frequency of encounters with others and thus opportunities for mixing. The recent evidence that many humans carry genomic segments that can be traced back to Neanderthals and Denisovans (a second archaic human group) is evidence of such genetic exchange even 30,000 or more years ago with then-existing archaic human groups (Pääbo 2014).

The history that I have outlined above has implications for the natural selection and adaptation that have surely shaped our genomes. First, the major adaptive events that led to the emergence of *H. sapiens* are not the subject of debate. For matters of race, it does not even matter what happened to our young species in all of the last 150,000 years, or the emergence of modern humans while in Africa. What does matter are the adaptations in the past 50,000 years that have led to our spread across all continents and the genetic differences between us since then. This span of 2000 generations is long enough for specific adaptations to have occurred but not for very many such adaptations. The reasons, albeit technical, essentially depend on the following argument. Every adaptation occurs through some individuals possessing a beneficial mutation that, while bringing them an advantage (larger numbers of surviving offspring), is a relative disadvantage to those who do not carry that beneficial variant. This is a precarious gift because early in the evolution of this mutation most individuals are at a disadvantage and even those with an advantage might not realize their benefit. The English geneticist J.B.S. Haldane argued that this cost of selection was high enough that many beneficial mutations cannot arise at the same time (Haldane 1957). Of course, adaptations do occur as is evidenced by the striking examples of lactose persistence (selected after dairy farming arose) or skin pigmentation (selected for in response to solar radiation) (Quintana-Murci and Barreiro 2010). To me, even more striking is the repeated birth of the sickle cell mutation at the same DNA site in Africa as a protective response to malaria (Wainscoat et al. 1983). Nevertheless, these examples of adaptation are rare and there are likely only a handful of such examples in genomes across humanity (Hernandez et al. 2011). Positive selection and adaptations must have occurred, but their mechanisms likely are not through single genes but across many genes affecting the same trait, as is the case for height (Turchin et al. 2012). If that is the common scenario, then even strong selection on one gene among the many that affect a trait is going to be trivial unless we are speaking of very long evolutionary times or adaptive changes that occurred in our shared history.

The thesis that human groups substantially differ in most traits that are deeply rooted in simple genetics, and the result of recent adaptation, is fanciful. Human groups do differ from one another in many ways and the reasons are more likely to be nongenetic than genetic. Of those that are genetic, their composition is likely owing to many genes (hundreds to thousands),

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

adaptations at many of them unlikely to be sustainable by known genetic mechanisms in the time frame during which human differences must have accumulated. More than 45 years ago, Motoo Kimura contended that, broadly, most of molecular evolution is deleterious and doomed to extinction; of those that do survive, the vast majority are selectively neutral (Kimura 1968). Recent data suggest this to be amply true. A benefit of this theory is that it means that the vast majority of changes in our genomes occur at a constant rate and provide an excellent "molecular clock" to date specific events in our common and unique histories.

THE NEXT PHASE OF HUMAN GENETIC DIVERSITY STUDIES

Human genetic diversity is dynamic and its patterns have changed substantially over time and will change in the future. Each of us can trace our genomes back to Africa and the subsequent journey across 50,000 years with intermixing with many other peoples. We live, however, in very different times. There are increasing rates of meeting and mixing, including groups that may have been relatively isolated for a few thousand generations. There are also rapid cultural and social changes that make neither gene nor cultural isolation possible. All of this implies even greater admixture than ever before. Consequently, I suspect the study of individuals and their genomes will increase at the expense of studying populations. It is remarkable how many individuals choose to study their genomes purely to decipher their ancestry (www.23andme.com). These individual genomes will surely uncover their individual histories but also begin to add detail to our common genetic history (1000 Genomes Consortium 2012).

I suspect that the focus on race, caste, or tribe that we still see today will erode simply because fewer and fewer members of any group will have its hallmark features. What does continental ancestry mean when one is from more than one? It might survive, I suppose; after all, how many fans of Manchester United across the globe have ever even been to Manchester? For a while, our ability to decipher individual histories might be useful to test how strict or porous the concept of a "population" is. Genetics and evolutionary biology have held as fundamental the concept that a population is a real, stable genetic unit, a property that is discrete and survivable. The reality is that most populations are dynamic and fluid, neither real nor stable.

Human evolution has always been studied with respect to such populations defined by language, geography, or cultural and physical features. Consider instead what we could decipher if we could sample a million humans (say), without regard to who they were, across a virtual grid across the world; this would correspond to sampling one person every 57 square miles ($\sim$7.5 miles $\times$ 7.5 miles) across the land surface. (This grid sampling idea was mentioned to me by the late Allan Wilson sometime around 1988.) Assume as well that we would sequence their maternal and paternal genomes and ask them questions, such as where they were born, where their parents were born, which language(s) they spoke, and what group affiliation(s) they had. We could then specifically uncover not only all of the features of human evolution we know today, and revise them to greater accuracy, but also test whether any or all of the features we use for human classification are supported by their genes. These types of global surveys of diversity have been performed for other species and may provide the first objective description of ours, bereft of race and other labels. This does not vitiate any social or cultural ways of defining humans, but at least one can no longer claim a genetic basis for all group differences.

RACE-BASED AND INDIVIDUALIZED MEDICINE

One of the major contributions of genetics to medicine has been, beyond the identification of disease pathophysiology, the recognition that each disease is multicausal and that patients with a single disease label may have conditions that arise from distinct molecular causes. This is well recognized for single gene disorders such as muscular dystrophy or even broad categories like prelingual deafness. Distinct molecular etiologies may require distinct therapies and man-

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

agement methods simply because physicians are trying to ameliorate different pathologies. These ideas have led to the concept of individualized medicine, that is, the tailoring of care to each patient depending on their genetic makeup and individual health circumstances (Childs 1999). This conceptual basis for care is a fundamental change in medicine that has long relied on the idea of a typological patient. Medicine's goal is now individualized care for the common chronic human diseases, a far more challenging task because the genetic and molecular bases for most chronic diseases remain unknown. We are making great progress but we have a long journey ahead before we understand the genetic, epigenetic, and environmental contributions to these disorders and which of these three may be the best route to intervention. That, of course, should not prevent us from individualizing care as best as we can with the knowledge we already have.

One of the major areas for individualized medicine is cancer therapy in which molecular diagnosis of germline mutations has been prevalent for more than two decades, and the genetic profile of the tumor has directed aspects of treatment. Increasingly, genome sequencing is being used to profile tumors broadly to direct treatment (Vogelstein et al. 2013). Individualizing risks to specific cancer subtypes and tumor profiling are expected to become routine aspects of cancer treatment particularly because personalized immune-modulation therapy is also on the horizon (Pardoll 2012). It is interesting to reflect that so much progress has been made in cancer treatment with greater discussion of personalized medicine rather than through the lens of race, despite cancer epidemiology data that show differences by continental ancestry. Differences in cancer incidence and prevalence by ancestry or ethnicity or community are well known. In fact, significant differences by geography, in the United States down to the county level, are also well known, suggesting major environmental etiologies as well. These differences, which lie beyond genes, need to be addressed simultaneously. The tussle lies in which genes and environments will we emphasize when both are responsible?

There is, of course, a long history of the study of variation in the incidence and prevalence of any disease by race, continental ancestry, and ethnicity. There is no doubt that many disorders show persistent and consistent differences and lead to great health disparities around the world (Murray et al. 2013). Genes do contribute to some part of it, probably no more than half, based on heritability studies, but environmental, epigenetic, and chance effects contribute significantly as well. Thus, equating all differences to genes is neither correct nor wise. As the essays by Richard Cooper and Troy Duster eloquently argue, the nongenetic factors in human disease, equally if not more importantly, affect our genetic biology in fundamental ways, and in many cases merit direct interventions that can lead to reductions in disease prevalence. Treatment of elevated blood pressure to prevent hypertension and its associated damage to the heart and kidney is a cogent example. In the United States, African Americans have elevated rates of hypertension and its sequelae as compared with those with European ancestry. This consistent finding has led to the myth that Africans have a higher genetic predisposition to hypertension, a fact clearly refuted by studies of many African communities whose blood pressures are lower than those of many European communities (Cooper et al. 2005). Moreover, recent genetic investigations in very large samples clearly show that blood pressure susceptibility variants detected in European ancestry subjects are also susceptibility variants for African, Asian, and South Asian subjects (Ehret et al. 2011). How could it not be so? Blood pressure regulation is a crucial human physiological trait under homeostasis, probably modulated by hundreds of genes, genetic variation of which was probably chosen in the early days of human evolution. It is not surprising that all humans likely share such variants. This is not to argue that additional variants did not arise later, variants that are consequently expected to show geographic clustering, but they are expected to be fewer. In other words, a race-based approach to medicine is a poor proxy for the type of genetic understanding we need to allow advanced medical interventions, like those available for cancer. Genetics

 Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a023358

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

can be far more useful in identifying the molecular underpinnings of human disease and treatment differences by focusing on all of humanity and including our diversity, not ignoring it (see Lu et al. 2014). Population differences do exist, but genes are not the sole agent for these differences, and nongenetic factors may have greater potency (Kahn 2013). An important corollary is that we should not speak of nature and nurture generally but take their respective roles on a case-by-case basis.

The challenge for understanding complex traits is thus considerable. Defining the roles of specific genes in the common chronic disorders of our time will lead to a much improved understanding of how and why a disorder develops (pathophysiology) and thus lead to improved therapies. And this progress will critically depend, I believe, on parallel progress in our understanding of how environmental and epigenetic factors impact our biology. These are the other sides of the genetic coin and there is no intellectual solution to one without the other. One of my mentors, Ching Chun (CC) Li, once remarked, "We are geneticists, not hereditarians; of course, the environment is important." As a plant breeder in his native China he had seen both genetically selected crops wither in a drought and the fantastical bogus claims of improved yields by Trofim Lysenko. We have to get the genetics content right this time. It is a fundamental biological challenge and an exciting one. It is even more exciting to figure out how the outside (environment) affects the inside (the genetic program) and, in parallel, how and why we as a species are so much more diverse in our cultures than in our genes. Vive la différence!

ACKNOWLEDGMENTS

I am indebted to Richard Cooper, John Novembre, and Richard Sever for critical comments on this paper.

REFERENCES

*Reference is also in this collection.

1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491:** 55–65.

Blumenbach JF. 1775. *De generis humani varietate nativa [On the natural varieties of mankind]*. University of Göttingen, Germany.

Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. *Nature* **325:** 31–36.

Cavalli-Sforza LL, Edwards AWF. 1964. Analysis of human evolution. *Proc 11th Int Cong Hum Genet* **2:** 923–933.

Cavalli-Sforza LL, Feldman MW. 1981. *Cultural transmission and evolution: A quantitative approach*. Princeton University Press, Princeton, NJ.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. *The history and geography of human genes*. Princeton University Press, Princeton, NJ.

Chakravarti A. 2010. *Principia genetica*: Our future science. *Am J Hum Genet* **86:** 302–308.

Childs B. 1999. *Genetic medicine: A logic of disease*. Johns Hopkins University Press, Baltimore.

* Cooper RS. 2013. Race in biological and biomedical research. *Cold Spring Harb Perspect Med* **3:** a008573.

Cooper RS, Wolf-Maier K, Adeyemo A, Banegas JR, Forrester T, Giampaoli S, Joffres M, Kasterinen M, Primastesta P, Stegmayr B, et al. 2005. An international comparative study of blood pressure in populations of European vs. African descent. *BMC Med* **3:** 2.

Darwin C. 1871. *The descent of man and selection in relation to sex*. John Murray, London.

deGrasse Tyson N. 2014. *The Colbert Report*, March 10, 2014.

* Duster T. 2014. Social diversity in humans: Implications and hidden consequences for biological research. *Cold Spring Harb Perspect Biol* **6:** a008482.

Edwards AWF, Cavalli-Sforza LL. 1964. Reconstruction of evolutionary trees. In *Phenetic and phylogenetic classification* (ed. Heywood VE, McNeill J), pp. 67–76. The Systematics Association, London.

Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, Hwang S-J, et al. 2011. Common polymorphisms impacting blood pressure and cardiovascular disease in diverse populations highlight novel biological pathways. *Nature* **478:** 103–109.

Fraser S. 1995. *The Bell curve wars: Race, intelligence, and the future of America*. Basic Books, New York.

Fredrickson GM. 2002. *Racism: A short history*. Princeton University Press, Princeton, NJ.

Gluckman P, Beedle A, Hanson M. 2009. *Principles of evolutionary medicine*. Oxford University Press, Oxford.

* Gomez F, Hirbo J, Tishkoff SA. 2014. Genetic variation and adaptation in Africa: Implications for human evolution and disease. *Cold Spring Harb Perspect Biol* **6:** a008524.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* **5:** e1000695.

Haldane JBS. 1957. The cost of natural selection. *J Genet* **55:** 511–524.

Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, 1000 Genomes Project, Sella G, Przeworski M. 2011. Classic selective sweeps were rare in recent human evolution. *Science* **331:** 920–924.

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.

This is a free sample of content from Human Variation.
Click here for more information on how to buy the book.

Herrnstein RJ, Murray C. 1994. *The Bell curve: Intelligence and class structure in American life*. Free Press, New York.

Jirtle RL, Skinner MK. 2007. Environmental epigenomics and disease susceptibility. *Nat Rev Genet* **8:** 253–262.

Kahn J. 2013. *Race in a bottle*. Columbia University Press, New York.

Kays JL, Hurley RA, Taber KH. 2012. The dynamic brain: Neuroplasticity and mental health. *J Neuropsych Clin Neurosci* **24:** 118–124.

Kimura M. 1968. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.

Klein RG. 1989. *The human career: Human biological and cultural origins*. University of Chicago Press, Chicago.

Koenig BA, Lee SS-J, Richardson SS. 2008. *Revisiting race in a genomic age*. Rutgers University Press, Piscataway, NJ.

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, et al. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467:** 832–838.

Lewontin R. 1972. The apportionment of human diversity. *Evol Biol* **6:** 391–398.

* Lu YF, Goldstein DB, Angrist M, Cavalleri G. 2014. Personalized medicine and human genetic diversity. *Cold Spring Harb Perspect Med* doi: 10.1101/cshperspect.a008581.

* Majumder PP, Basu A. 2014. A genomic view of the peopling and population structure of India. *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a008540.

Mayr E. 2002. The biology of race and the concept of equality. *Daedalus* **131:** 89–94.

Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, et al. 2013. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380:** 2197–2223.

Orel V. 1996. *Gregor Mendel: The first geneticist*. Oxford University Press, Oxford.

Pääbo S. 2014. The human condition—A molecular approach. *Cell* **157:** 216–226.

Pardoll DM. 2012. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* **12:** 252–264.

Provine WB. 1971. *The origins of theoretical population genetics*. University of Chicago Press, Chicago.

Quintana-Murci L, Barreiro LB. 2010. The role played by natural selection on Mendelian traits in humans. *Ann NY Acad Sci* **1214:** 1–17.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. *Nature* **461:** 489–494.

* Ruiz Linares A. 2014. How genes have illuminated the history of early Americans and Latino Americans. *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a008557.

Smithsonian's National Museum of Natural History. 2013. Panel discussion on "ancestry and health." National Museum of African American History and Culture, National Human Genome Research Institute, September 12, 2013.

* Thapar R. 2014. Can genetics help us understand Indian social history? *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a008599.

Turchin MC, Chiang CW, Palmer CD, Sankararaman S, Reich D, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, Hirschhorn JN. 2012. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat Genet* **44:** 1015–1019.

* Veeramah KR, Novembre J. 2014. Demographic events and evolutionary forces shaping European genetic diversity. *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a008516.

Vinkhuyzen AA, Wray NR, Yang J, Goddard ME, Visscher PM. 2013. Estimation and partition of heritability in human populations using whole-genome analysis methods. *Annu Rev Genet* **47:** 75–95.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. 2013. Cancer genome landscapes. *Science* **339:** 1546–1558.

Wade N. 2014. *A troublesome inheritance: Genes, race and human history*. Penguin, New York.

Wainscoat JS, Bell JI, Thein SL, Higgs DR, Sarjeant GR, Peto TE, Weatherall DJ. 1983. Multiple origins of the sickle mutation: Evidence from βS globin gene cluster polymorphisms. *Mol Biol Med* **1:** 191–197.

* Weiss KM, Lambert BW. 2014. What type of person are you? Old-fashioned thinking even in modern science. *Cold Spring Harb Perspect Biol* **6:** a021238.

Witkowski J, Inglis J (eds.). 2008. *Davenport's dream: 21st century reflections on heredity and eugenics*. Cold Spring Harbor Laboratroy Press, Cold Spring Harbor, NY.

© 2014 by Cold Spring Harbor Laboratory Press. All rights reserved.