

# WHAT MAKES US HUMAN?

Comparisons of the genomes of humans and chimpanzees are revealing those rare stretches of DNA that are ours alone

By Katherine S. Pollard

## KEY CONCEPTS

- Chimpanzees are the closest living relatives of humans and share nearly 99 percent of our DNA.
- Efforts to identify those regions of the human genome that have changed the most since chimps and humans diverged from a common ancestor have helped pinpoint the DNA sequences that make us human.
- The findings have also provided vital insights into how chimps and humans can differ so profoundly, despite having nearly identical DNA blueprints.

—The Editors

Six years ago I jumped at an opportunity to join the international team that was identifying the sequence of DNA bases, or “letters,” in the genome of the common chimpanzee (*Pan troglodytes*). As a biostatistician with a long-standing interest in human origins, I was eager to line up the human DNA sequence next to that of our closest living relative and take stock. A humbling truth emerged: our DNA blueprints are nearly 99 percent identical to theirs. That is, of the three billion letters that make up the human genome, only 15 million of them—less than 1 percent—have changed in the six million years or so since the human and chimp lineages diverged.

Evolutionary theory holds that the vast majority of these changes had little or no effect on our biology. But somewhere among those roughly 15 million bases lay the differences that made us human. I was determined to find them. Since then, I and others have made tantalizing progress in identifying a number of DNA sequences that set us apart from chimps.

### An Early Surprise

Despite accounting for just a small percentage of the human genome, millions of bases are still a vast territory to search. To facilitate the hunt, I wrote a computer program that would scan the

human genome for the pieces of DNA that have changed the most since humans and chimps split from a common ancestor. Because most random genetic mutations neither benefit nor harm an organism, they accumulate at a steady rate that reflects the amount of time that has passed since two living species had a common forebear (this rate of change is often spoken of as the “ticking of the molecular clock”). Acceleration in that rate of change in some part of the genome, in contrast, is a hallmark of positive selection, in which mutations that help an organism survive and reproduce are more likely to be passed on to future generations. In other words, those parts of the code that have undergone the most modification since the chimp-human split are the sequences that most likely shaped humankind.

In November 2004, after months of debugging and optimizing my program to run on a massive computer cluster at the University of California, Santa Cruz, I finally ended up with a file that contained a ranked list of these rapidly evolving sequences. With my mentor David Haussler leaning over my shoulder, I looked at the top hit, a stretch of 118 bases that together became known as human accelerated region 1 (HAR1). Using the U.C. Santa Cruz genome browser, a visualization tool that annotates the





**THE 1 PERCENT DIFFERENCE:** Humans are distinct from chimpanzees in a number of important respects, despite sharing nearly 99 percent of their DNA. New analyses are revealing which parts of the genome set our species apart.

human genome with information from public databases, I zoomed in on HAR1. The browser showed the HAR1 sequences of a human, chimp, mouse, rat and chicken—all of the vertebrate species whose genomes had been decoded by then. It also revealed that previous large-scale screening experiments had detected HAR1 activity in two samples of human brain cells, although no scientist had named or studied the sequence yet. We yelled, “Awesome!” in unison when we saw that HAR1 might be part of a gene new to science that is active in the brain.

We had hit the jackpot. The human brain is well known to differ considerably from the chimpanzee brain in terms of size, organization and complexity, among other traits. Yet the developmental and evolutionary mechanisms underlying the characteristics that set the human brain apart are poorly understood. HAR1 had the potential to illuminate this most mysterious aspect of human biology.

We spent the next year finding out all we could about the evolutionary history of HAR1 by comparing this region of the genome in vari-

ous species, including 12 more vertebrates that were sequenced during that time. It turns out that until humans came along, HAR1 evolved extremely slowly. In chickens and chimps—whose lineages diverged some 300 million years ago—only two of the 118 bases differ, compared with 18 differences between humans and chimps, whose lineages diverged far more recently. The fact that HAR1 was essentially frozen in time through hundreds of millions of years indicates that it does something very important; that it then underwent abrupt revision in humans suggests that this function was significantly modified in our lineage.

A critical clue to the function of HAR1 in the brain emerged in 2005, after my collaborator Pierre Vanderhaeghen of the Free University of Brussels obtained a vial of HAR1 copies from our laboratory during a visit to Santa Cruz. He used these DNA sequences to design a fluorescent molecular tag that would light up when HAR1 was activated in living cells—that is, copied from DNA into RNA. When typical genes are switched on in a cell, the cell first



makes a mobile messenger RNA copy and then uses the RNA as a template for synthesizing some needed protein. The labeling revealed that HAR1 is active in a type of neuron that plays a key role in the pattern and layout of the developing cerebral cortex, the wrinkled outermost brain layer. When things go wrong in these neurons, the result may be a severe, often deadly, congenital disorder known as lissencephaly ("smooth brain"), in which the cortex lacks its characteristic folds and exhibits a markedly reduced surface area. Malfunctions in these same neurons are also linked to the onset of schizophrenia in adulthood.

HAR1 is thus active at the right time and place to be instrumental in the formation of a healthy cortex. (Other evidence suggests that it may additionally play a role in sperm production.) But exactly how this piece of the genetic code affects cortex development is a mystery my colleagues and I are still trying to solve. We are eager to do so: HAR1's recent burst of substitutions may have altered our brains significantly.

Beyond having a remarkable evolutionary history, HAR1 is special because it does not encode a protein. For decades, molecular biology research focused almost exclusively on genes that specify proteins, the basic building blocks of cells. But thanks to the Human Genome Project, which sequenced our own genome, scientists now know that protein-coding genes make up just 1.5 percent of our DNA. The other 98.5 percent—sometimes referred to as junk DNA—contains regulatory sequences that tell other genes when to turn on and off and genes encoding RNA that does not get translated into a protein, as well as a lot of DNA having purposes scientists are only beginning to understand.

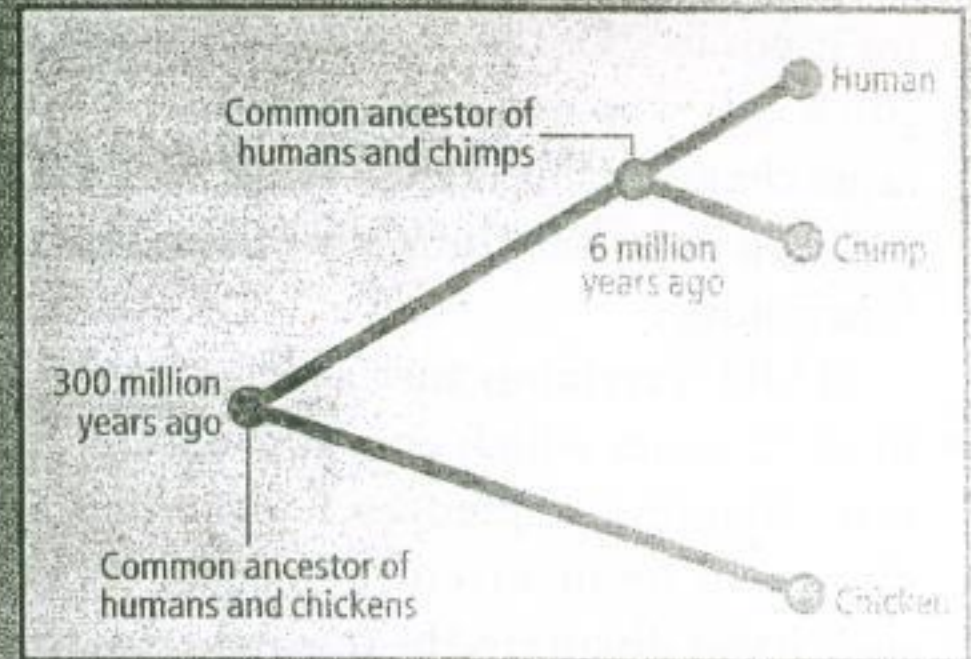
Based on patterns in the HAR1 sequence, we predicted that HAR1 encodes RNA—a hunch that Sofie Salama, Haller Igel and Manuel Ares, all at U.C. Santa Cruz, subsequently confirmed in 2006 through lab experiments. In fact, it turns out that human HAR1 resides in two overlapping genes. The shared HAR1 sequence gives rise to an entirely new type of RNA structure, adding to the six known classes of RNA genes. These six major groups encompass more than 1,000 different families of RNA genes, each one distinguished by the structure and function of the encoded RNA in the cell. HAR1 is also the first documented example of an RNA-encoding sequence that appears to have undergone positive selection.

It might seem surprising that no one paid at-

[EXPERIMENT]

# SCANNING THE GENOME

To find the parts of our genome that make us human, the author wrote a computer program that searched for the DNA sequences that have changed the most since humans and chimpanzees diverged from their last common ancestor. Topping the list was a 118-letter snippet of code known as human accelerated region 1 (HAR1). This region of the genome changed very little for most of vertebrate evolution, with chimp and chicken sequences differing by just two letters. Human and chimp HAR1s, however, differ by 18 letters, suggesting that HAR1 acquired an important new function in humans.



T	G	A	A	C	G	G	A	G	G	A	G	A	C	G	T	T	A	C	
A	G	C	A	A	C	G	T	G	T	C	A	G	C	T	G	A	A	A	T
G	A	T	G	G	C	G	T	A	G	A	C	G	C	A	C	G	T	C	
A	G	C	G	G	C	G	G	A	A	A	T	G	G	T	T	T	C	T	A
T	C	A	A	A	A	T	G	A	A	A	G	T	G	T	T	T	A	G	A
G	A	T	T	T	C	C	T	C	A	A	G	T	T	T	C	A			

Changes in human sequence relative to that of the chimp

T	G	A	A	A	T	G	G	A	G	G	A	G	A	A	A	T	T	A	C
A	G	C	A	A	T	T	T	A	T	C	A	A	C	T	G	A	A	A	T
T	A	T	A	G	G	T	G	T	A	G	A	C	A	C	A	T	G	T	C
A	G	C	A	G	T	G	G	A	A	A	T	A	G	T	T	T	C	T	A
T	C	A	A	A	A	T	T	A	A	A	G	T	A	T	T	T	A	G	A
G	A	T	T	T	C	C	T	C	A	A	A	T	T	T	C	A			

Changes in chimp sequence relative to that of the chicken

T	G	A	A	A	T	G	G	A	G	G	A	G	A	A	A	T	T	A	C
A	G	C	A	A	T	T	T	A	T	C	A	A	C	T	G	A	A	A	T
T	A	T	A	G	G	T	G	T	A	G	A	C	A	C	A	T	G	T	C
A	G	C	A	G	T	A	G	A	A	A	C	A	G	T	T	T	C	T	A
T	C	A	A	A	A	T	T	A	A	A	G	T	A	T	T	T	A	G	A
G	A	T	T	T	C	C	T	C	A	A	A	T	T	T	C	A			

attention to these amazing 118 bases of the human genome earlier. But in the absence of technology for readily comparing whole genomes, researchers had no way of knowing that HAR1 was more than just another piece of junk DNA.

### Language Clues

Whole-genome comparisons in other species have also provided another crucial insight into why humans and chimps can be so different despite being much alike in their genomes. In recent years the genomes of thousands of species (mostly microbes) have been sequenced. It turns out that where DNA substitutions occur



in the genome—rather than how many changes arise overall—can matter a great deal. In other words, you do not need to change very much of the genome to make a new species. The way to evolve a human from a chimp-human ancestor is not to speed the ticking of the molecular clock as a whole. Rather the secret is to have rapid change occur in sites where those changes make an important difference in an organism's functioning.

HAR1 is certainly such a place. So, too, is the *FOXP2* gene, which contains another of the fast-changing sequences I identified and is known to be involved in speech. Its role in speech was discovered by researchers at the University of Oxford in England, who reported in 2001 that people with mutations in the gene are unable to make certain subtle, high-speed facial movements needed for normal human speech, even though they possess the cognitive ability to process language. The typical human sequence displays several differences from the chimp's: two base substitutions that altered its protein product and many other substitutions that may have led to shifts affecting how, when and where the protein is used in the human body.

A recent finding has shed some light on when the speech-enabling version of *FOXP2* appeared in hominids: in 2007 scientists at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, sequenced *FOXP2* extracted from a Neandertal fossil and found that these extinct humans had the modern human version of the gene, perhaps permitting them to enunciate as we do. Current estimates for when the Neandertal and modern human lineages split suggest that the new form of *FOXP2* must have emerged at least half a million years ago. Most of what distinguishes human language from vocal communication in other species, however, comes not from physical means but cognitive ability, which is often correlated with brain size. Primates generally have a larger brain than would be expected from their body size. But human brain volume has more than tripled since the chimp-human ancestor—a growth spurt that genetics researchers have only begun to unravel.

One of the best-studied examples of a gene linked to brain size in humans and other animals is *ASPM*. Genetic studies of people with a condition known as microcephaly, in which the brain is reduced by up to 70 percent, uncovered the role of *ASPM* and three other genes—*MCPH1*, *CDK5RAP2* and *CENPJ*—in con-

**You do not need to change very much of the genome to make a new species.**

trolling brain size. More recently, researchers at the University of Chicago and the University of Michigan at Ann Arbor have shown that *ASPM* experienced several bursts of change over the course of primate evolution, a pattern indicative of positive selection. At least one of these bursts occurred in the human lineage since it diverged from that of chimps and thus was potentially instrumental in the evolution of our large brains.

Other parts of the genome may have influenced the metamorphosis of the human brain less directly. The computer scan that identified HAR1 also found 201 other human accelerated regions, most of which do not encode proteins or even RNA. (A related study conducted at the Wellcome Trust Sanger Institute in Cambridge, England, detected many of the same HARs.) Instead they appear to be regulatory sequences that tell nearby genes when to turn on and off. Amazingly, more than half of the genes located near HARs are involved in brain development and function. And, as is true of *FOXP2*, the products of many of these genes go on to regulate other genes. Thus, even though HARs make up a minute portion of the genome, changes in these regions could have profoundly altered the human brain by influencing the activity of whole networks of genes.

#### [THE AUTHOR]



**Katherine S. Pollard** is a biostatistician at the University of California, San Francisco. In 2003, after completing a Ph.D. and postdoctoral research at U.C. Berkeley, she began a comparative genomics fellowship at U.C. Santa Cruz, during which she participated in the sequencing of the chimpanzee genome. Pollard used this sequence to identify the fastest-evolving regions of the human genome. A recipient in 2008 of a Sloan Research Fellowship in Computational and Evolutionary Molecular Biology, she recently began studying the evolution of microbes that live in the human body.

#### Beyond the Brain

Although much genetic research has focused on elucidating the evolution of our sophisticated brain, investigators have also been piecing together how other unique aspects of the human body came to be. HAR2, a gene regulatory region and the second most accelerated site on my list, is a case in point. In 2008 researchers at Lawrence Berkeley National Laboratory showed that specific base differences in the human version of HAR2 (also known as HACNS1), relative to the version in nonhuman primates, allow this DNA sequence to drive gene activity in the wrist and thumb during fetal development, whereas the ancestral version in other primates cannot. This finding is particularly provocative because it could underpin morphological changes in the human hand that permitted the dexterity needed to manufacture and use complex tools.

Aside from undergoing changes in form, our ancestors also underwent behavioral and physiological shifts that helped them adapt to altered circumstances and migrate into new environments. For example, the conquest of fire more than a million years ago and the agricultural



[FINDINGS]

## DISTINCTIVE DNA

Efforts to uncover uniquely human DNA have yielded a number of sequences that are distinctive in humans as compared with chimpanzees. A partial list of these sequences and some of their functions—follows below.

### SEQUENCE: HAR1

**What it does:** Active in the brain, may be necessary for development of the cerebral cortex, which is especially large in humans. Possibly also involved in sperm production.

### SEQUENCE: FOXP2

**What it does:** Facilitates formation of words by the mouth, enabling modern human speech.

### SEQUENCE: AMY1

**What it does:** Facilitates digestion of starch, which may have enabled early humans to exploit novel foods.

### SEQUENCE: ASPM

**What it does:** Controls brain size, which has more than tripled over the course of human evolution.

### SEQUENCE: LCT

**What it does:** Permits digestion of milk sugar in adulthood, allowing people to make milk from domesticated animals a dietary staple.

### SEQUENCE: HAR2

**What it does:** Drives gene activity in the wrist and thumb during development, an activity that may have given the hand enough dexterity to make and use complex tools.

revolution about 10,000 years ago made foods high in starch more accessible. But cultural shifts alone were not sufficient to exploit these calorie-rich comestibles. Our predecessors had to adapt genetically to them.

Changes in the gene *AMY1*, which encodes salivary amylase, an enzyme involved in digesting starch, constitute one well-known adaptation of this kind. The mammalian genome contains multiple copies of this gene, with the number of copies varying between species and even between individual humans. But overall, compared with other primates, humans have an especially large number of *AMY1* copies. In 2007 geneticists at Arizona State University showed that individuals carrying more copies of *AMY1* have more amylase in their saliva, thereby allowing them to digest more starch. The evolution of *AMY1* thus appears to involve both the number of copies of the gene and the specific changes in its DNA sequence.

Another famous example of dietary adaptation involves the gene for lactase (*LCT*), an enzyme that allows mammals to digest the carbohydrate lactose, also known as milk sugar. In most species, only nursing infants can process lactose. But around 9,000 years ago—very recently, in evolutionary terms—changes in the human genome produced versions of *LCT* that allowed adults to digest lactose. Modified *LCT* evolved independently in European and African populations, enabling carriers to digest milk from domesticated animals. Today adult descendants of these ancient herders are much more likely to tolerate lactose in their diets than are adults from other parts of the world, including Asia and Latin America, many of whom are lactose-intolerant as a result of having the ancestral primate version of the gene.

*LCT* is not the only gene known to be evolving in humans right now. The chimp genome project identified 15 others in the process of shifting away from a version that was perfectly normal in our ape ancestors and that works fine in other mammals but, in that old form, is associated with diseases such as Alzheimer's and cancer in modern humans. Several of these disorders afflict humans alone or occur at higher rates in humans than in other primates. Scientists are currently researching the functions of the genes involved and are attempting to establish why the ancestral versions of these genes became maladaptive in us. These studies could help medical practitioners identify those patients who have a higher chance of getting one



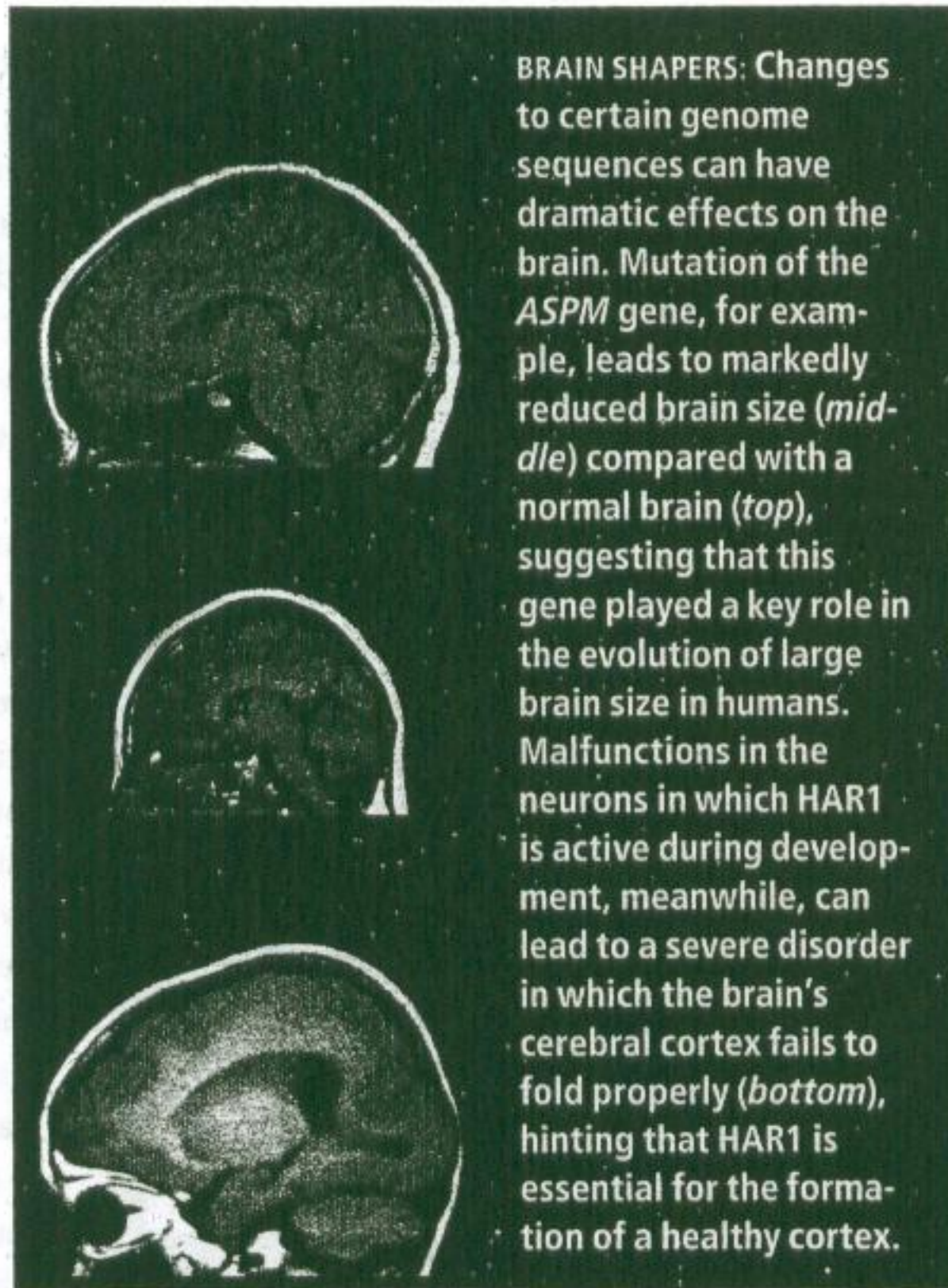
of these life-threatening diseases, in hopes of helping them stave off illness. The studies may also help researchers identify and develop new treatments.

### With the Good Comes the Bad

Battling disease so we can pass our genes along to future generations has been a constant refrain in the evolution of humans, as in all species. Nowhere is this struggle more evident than in the immune system. When researchers examine the human genome for evidence of positive selection, the top candidates are frequently involved in immunity. It is not surprising that evolution tinkers so much with these genes: in the absence of antibiotics and vaccines, the most likely obstacle to individuals passing along their genes would probably be a life-threatening infection that strikes before the end of their childbearing years. Further accelerating the evolution of the immune system is the constant adaptation of pathogens to our defenses, leading to an evolutionary arms race between microbes and hosts.

Records of these struggles are left in our DNA. This is particularly true for retroviruses, such as HIV, that survive and propagate by inserting their genetic material into our genomes. Human DNA is littered with copies of these short retroviral genomes, many from viruses that caused diseases millions of years ago and that may no longer circulate. Over time the retroviral sequences accumulate random mutations just as any other sequence does, so that the different copies are similar but not identical. By examining the amount of divergence among these copies, researchers can use molecular clock techniques to date the original retroviral infection. The scars of these ancient infections are also visible in the host immune system genes that constantly adapt to fight the ever evolving retroviruses.

PtERV1 is one such relic virus. In modern humans, a protein called TRIM5 $\alpha$  works to prevent PtERV1 and related retroviruses from replicating. Genetic evidence suggests that a PtERV1 epidemic plagued ancient chimpanzees, gorillas and humans living in Africa about four million years ago. To figure out how different primates re-



**BRAIN SHAPERS:** Changes to certain genome sequences can have dramatic effects on the brain. Mutation of the *ASPM* gene, for example, leads to markedly reduced brain size (*middle*) compared with a normal brain (*top*), suggesting that this gene played a key role in the evolution of large brain size in humans. Malfunctions in the neurons in which *HAR1* is active during development, meanwhile, can lead to a severe disorder in which the brain's cerebral cortex fails to fold properly (*bottom*), hinting that *HAR1* is essential for the formation of a healthy cortex.

sponded to PtERV1, in 2007 researchers at the Fred Hutchinson Cancer Research Center in Seattle used the many randomly mutated copies of PtERV1 in the chimpanzee genome to reconstruct the original PtERV1 sequence and re-create this ancient retrovirus. They then performed experiments to see how well the human and great ape versions of the TRIM5 $\alpha$  gene could restrict the activity of the resurrected PtERV1 virus. Their results indicate that a single change in human TRIM5 $\alpha$  most likely enabled our ancestors to fight PtERV1 infection more effectively than our primate cousins could. (Additional changes in human TRIM5 $\alpha$  may have evolved in response to a related retrovirus.) Other primates have their own sets of changes in TRIM5 $\alpha$ , probably reflecting retroviral battles that their predecessors won.

Defeating one type of retrovirus does not necessarily guarantee continued success against others, however. Although changes in human TRIM5 $\alpha$  may have helped us survive PtERV1, these same shifts make it much harder for us to fight HIV. This finding is helping researchers to understand why HIV infection leads to AIDS in humans but not in nonhuman primates. Clearly, evolution can take one step forward and two steps back. Sometimes scientific research feels the same way. We have identified many exciting candidates for explaining the genetic basis of distinctive human traits. In most cases, though, we know only the basics about the function of these genome sequences. The gaps in our knowledge are especially large for regions such as HAR1 and HAR2 that do not encode proteins.

These rapidly evolving, uniquely human sequences do point to a way forward. The story of what made us human is probably not going to focus on changes in our protein building blocks but rather on how evolution assembled these blocks in new ways by changing when and where in the body different genes turn on and off. Experimental and computational studies now under way in thousands of labs around the world promise to elucidate what is going on in the 98.5 percent of our genome that does not code for proteins. It is looking less and less like junk every day. ■

### ➔ MORE TO EXPLORE

**Mapping Human History: Discovering the Past through Our Genes.** Steve Olson. Houghton Mifflin, 2002.

**The Ancestor's Tale: A Pilgrimage to the Dawn of Evolution.** Richard Dawkins. Houghton Mifflin, 2004.

**Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome.** The Chimpanzee Sequencing and Analysis Consortium in *Nature*, Vol. 437, pages 69–87; September 1, 2005.

University of California, Santa Cruz, Genome Bioinformatics Web site: <http://genome.ucsc.edu>

COURTESY OF GANESHWARAN H. MOCHIDA, C. GEOFFREY WOODS AND CHRISTOPHER A. WALSH (top and middle); COURTESY OF GAVIN USTUEN (University of Cincinnati) (bottom)