

The Genome's Black Box

>> By Maya Pines

At the waistline of our chromosomes, the mysterious centromere holds a key to cancer and birth defects—and may reveal a new code in DNA.

IT'S THE ULTIMATE BLACK BOX OF OUR GENOME," SAYS HUNTINGTON E. Willard, "and it doesn't play according to the rules."

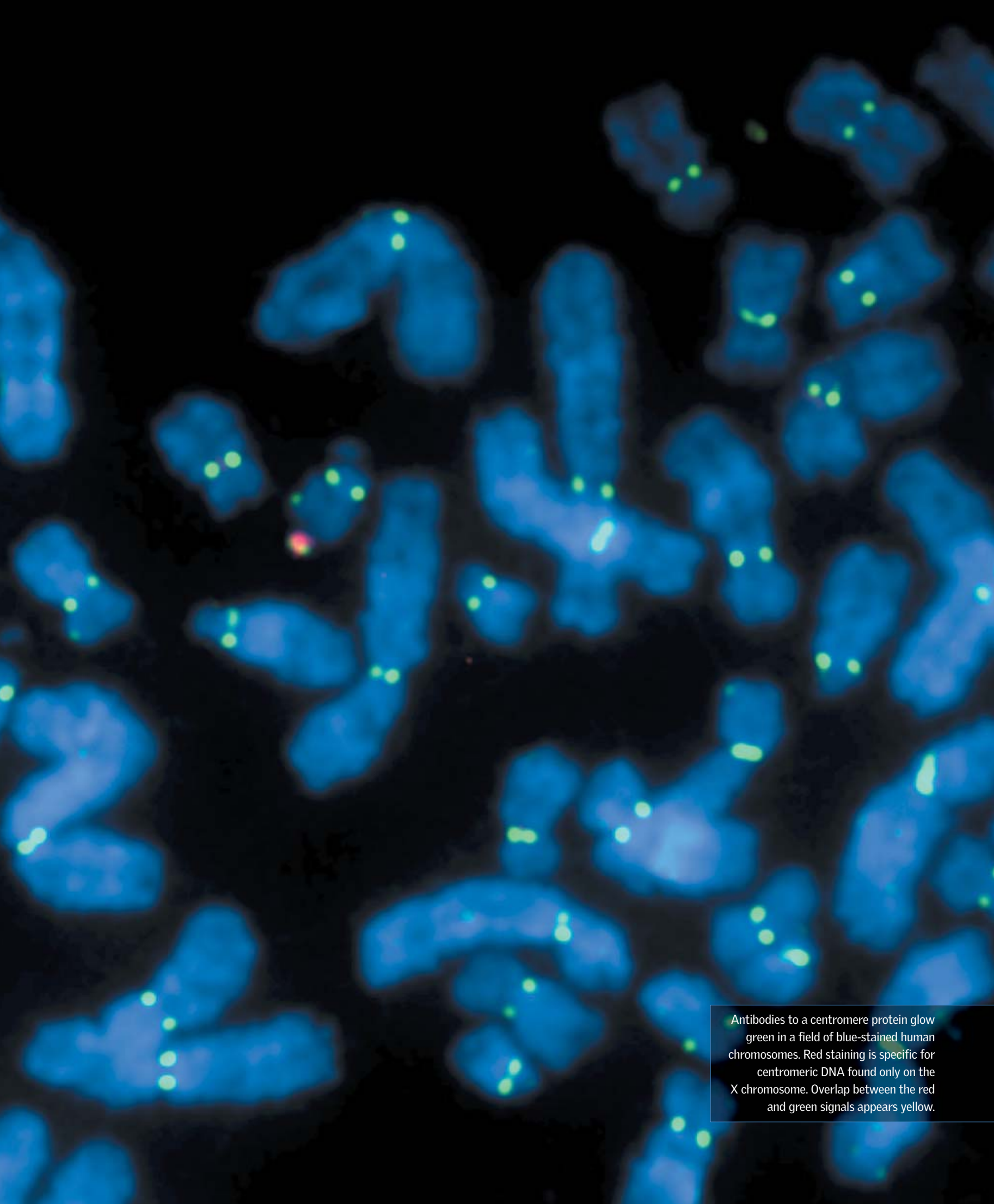
Willard, who heads the Institute for Genome Sciences and Policy at Duke University and is a member of HHMI's Scientific Review Board, spends much of his time thinking about this peculiar piece of DNA, called a centromere. Sitting at the waistline of our chromosomes, centromeres have a vital job—to direct the shuffling of chromosomes during cell division. And although they are generally reliable, they do make genetic mistakes—"perhaps once in a hundred cell divisions," says Willard.

These mistakes are responsible for "a majority of human disorders," he declares. "Look at the high frequency of birth defects that result from an abnormal number of chromosomes—1 in 700 births for Down syndrome, 1 in 500 births for Klinefelter's [in which males are born with an additional X chromosome]. And look at the number of people with cancer—1 in 3 of us!" The evil effects of cancer-causing genes are greatly speeded up by mistakes in the segregation of the chromosomes that carry these genes, he points out. "Though cancer researchers study specific genes that may lead to cancer, they often forget that mis-segregation of chromosomes is an underlying problem.

"Without a fully functional centromere, the genome that everyone is studying so hard wouldn't work, because the genes couldn't be appropriately partitioned during cell division," he says. "This would lead to gross imbalance, in which different cells had different numbers of each chromosome."

Human centromeres remain mysterious because their DNA, which is called alpha satellite DNA and makes up 3 to 4 percent of the entire human genome,

COURTESY OF WILLARD LAB.



Antibodies to a centromere protein glow green in a field of blue-stained human chromosomes. Red staining is specific for centromeric DNA found only on the X chromosome. Overlap between the red and green signals appears yellow.

differs from all the rest. It seems to consist of hundreds of thousands of repetitive segments whose exact order is difficult to determine—so difficult, in fact, that the scientists who sequenced the human genome never even attempted it. Even HHMI investigator David C. Page of the Massachusetts Institute of Technology, who recently sequenced the Y chromosome's masses of repetitive segments, stopped at its centromere. "We only sampled a few bits of DNA within the Y centromere and then gave up because we were unable to interpret it in a biologically interesting way," Page says.

As Willard emphasizes, "We have no idea how these segments operate. We just don't know the code." He hopes to find out, however. So do a growing number of other researchers, including Steven Henikoff, an HHMI investigator at the Fred Hutchinson Cancer Research Center in Seattle, Washington, who studies centromeres primarily in the fruit fly, and Daphne Preuss, an HHMI investigator at the University of Chicago, who works with the model plant *Arabidopsis thaliana*.

>> Violating a Basic Rule

The most intriguing thing about centromeres, Willard says, is that they seem to violate a basic rule of biology. Normally, all DNA sequences that code for important functions are "conserved," or repeated almost unchanged, from species to species throughout evolution. Nearly every mouse gene has a counterpart in humans, for instance, and some blocks of sequenced mouse DNA can barely be distinguished from human versions. Large numbers of yeast, worm, and fruit fly genes are also similar to their counterparts in the human genome. "Yet here's an absolutely critical *function* that's highly conserved in all animals—the segregation of chromosomes during cell division—but its DNA sequences are not conserved," Willard declares.

At the DNA level, the centromeres of mouse, fly, yeast, plant, and human look nothing like each other, he says. Even more surprising, worms thrive with not 1 but with 5 to 10 centromeres on each chromosome. "Nature has found a totally different way to deal with this important function," says Willard. "What's conserved is not the genomic sequence per se but what I'd call the 'centromere code,' and we don't understand that at all."

Willard became interested in centromeres in the early 1980s, at a time when only the yeast centromere—a tiny sequence of nonrepetitive DNA called "the magic sequence," only 125 bases long—had been analyzed. It was assumed that human centromeres would be roughly the same size. If so, they would be hard to find within the seemingly endless string of repetitive sequences in the middle section of human chromosomes.

"By luck," says Willard, "as I was studying some repetitive sequences of DNA from the human X chromosome, I came across a whole family of tandem sequences—171 base pairs of DNA that were repeated head to tail. Our lab cloned thousands of these alpha satellites. As a fairly naive researcher, I felt that this stuff must be doing something. That was quite heretical at the time, since most people considered that such sequences were 'junk.'"



"Without a fully functional centromere, the genome that everyone is studying so hard wouldn't work, because the genes couldn't be appropriately partitioned during cell division."

—HUNTINGTON WILLARD, *Institute for Genome Sciences and Policy, Duke University*

He began to wonder what these alpha satellites were doing around the human centromere. "Perhaps in our DNA there is no magic sequence that functions as a centromere," he speculated. "Perhaps it's the repetitive DNA itself that's doing it."

>> Help from Artificial Chromosomes

The best way to test this idea was to insert long stretches of alpha satellite into a chromosome and see whether the chromosome segregated normally. For this purpose, Willard set out to build a human artificial chromosome (HAC) from the ground up, using bits of human genetic material.

He knew that yeast artificial chromosomes (YACs—see sidebar) could be assembled from three component parts: telomeres (specialized sequences of DNA that are found at the ends of chromosomes and are conserved in almost all animal species); segments of DNA containing genes and the so-called origins of replication, which ensure that the synthesis of DNA is started correctly; and a yeast centromere, which

attaches the chromosome to a spindle of fibers that pulls the sister chromatids apart during cell division.

Could he use the same technique to make human chromosomes? The first two elements would be relatively easy to clone. The problem, of course, was the centromere. Willard made the “leap of faith” that millions of base pairs of alpha satellite DNA would do the job.

This plan got a jump-start in the mid-1990s when two postdoctoral fellows, John J. Harrington and Gil Van Bokkelen, joined Willard’s lab for the express purpose of building a HAC. The two young men were on a mission: They wanted to create safer and more reliable vehicles for human gene therapy, which attempts to slip missing genes into a cell’s nucleus to cure various diseases.

>> Better Gene Therapy?

Existing gene-therapy methods are both limited and risky. The modified viruses that are often used to carry inserts of DNA have space only for short pieces and cannot accept larger genes. Furthermore, any inserted DNA that lands in the middle of a preexisting human chromosome could cause mutations that might lead to cancer. On the other hand, DNA that remains separate from a chromosome is deprived of a working centromere and destined to disappear during cell division.

HACs would have enormous advantages, at least in theory. They would be able to carry large genes. And because they would remain independent of the cell’s preexisting chromosomes, they would not interfere with the original genes. They would simply go on reproducing within the cell nucleus, together with their imported genes.

With this exciting vision in mind, the scientists went to work. “It was a good marriage,” says Willard. “The two postdocs who brought us the gene-therapy angle needed a functioning centromere because that’s the only thing that would keep an artificial chromosome stable and segregating correctly inside a cell. And we knew that once we had a working artificial chromosome, we could take bite-sized pieces of it, one at a time, and really dissect them” to see how the parts worked.

It all happened according to plan. After several attempts, “we put in three types of DNA, and the chromosome self-assembled,” Willard declared. In April 1997, his team announced the creation of the first HAC. Since then, several other groups have succeeded in building versions of their own. These various HACs have functioned only in cultured cells in the lab, however. Scientists still need to find some way of delivering them safely into ordinary cells in a living organism.

Although successful gene therapy is still a long way off, artificial chromosomes will have many other uses, Willard believes. They will enable scientists to explore the function of specific genes, as well as the structure of chromosomes and the genome. They will illuminate the control of cell division. Most importantly, he hopes, they will bring clues to understanding the code that underlies centromeres and, perhaps, other kinds of so-called junk DNA.

>> What Junk DNA?

“Centromeres are the strongest case for the argument that junk DNA is not junk—that there must be some functions there,” Willard says. Biologists used to believe that all repetitive DNA that does not code for proteins—more than 50 percent of the human genome—was useless. Little by little, however, scientists have discovered the value of much of this “fluff,” as Willard prefers to call it. “It’s a more polite term,” he explains. “The real question is, Is this fluff there purely by accident, or is it the result of nature’s evolutionary testing? We shouldn’t put it down as junk just because we don’t understand the code yet.”

He points out that although human centromeres consist largely of alpha satellite, mice don’t have any alpha satellite at all in that region. “Yet they obviously have centromeres!” he says. “Their centromeres are made of something different, called minor satellite.

“I never had a moment like that...I knew it would fundamentally change everything you could do in plants.”

—DAPHNE PREUSS,
University of Chicago



YACs, HACs, MACs, PLACs, and FACs

Sometimes the best way to analyze a structure is to try to build it. This is what HHMI investigator Jack W. Szostak of the Massachusetts General Hospital in Boston decided to do 20 years ago, when he made the world's first yeast artificial chromosome (YAC) that looked—and acted—pretty much like a natural one. His goal was to dissect the structure of chromosomes.

Almost immediately, Maynard Olson, who is now director of the University of Washington Genome Center, saw the potential of YACs as carriers of large chunks of alien DNA that could be reproduced and

then mapped as part of the Human Genome Project. YACs soon became the workhorse of genome mapping in yeast and mammals.

“We then cottoned to this idea as an approach for human gene therapy—to introduce a therapeutic gene on its own private chromosome,” says Huntington Willard. His lab created the first human artificial chromosome (HAC) in 1997 with the help of thousands of DNA tandem repeats that mimicked a human centromere. And now he visualizes many more uses for HACs, quite unrelated to gene therapy (see main story).

Apparently, no self-respecting model organism can do without artificial chromo-

somes of its own any more. Because mouse centromeres are very different from human ones (although many mouse genes are almost identical to human genes), it is not so easy to create a mouse artificial chromosome (MAC). A MAC, however, is in the works.

Meanwhile, Daphne Preuss is making progress on a plant artificial chromosome but does not want to call it “PLAC”—that sounds unpleasant, she thinks. She plans to name it AtAC for the plant's full name, *Arabidopsis thaliana*. Not to be outdone, the fly research community is making its own plans, and soon the artificial chromosomes' bubbling alphabet soup may include another newcomer, FAC.

Dogs, cats, all animals beyond the primates have DNA sequences that are different from ours in their centromeres. Yet, they all have something in common—some code that is telling the cell how to use repetitive DNA to be a centromere.”

This is why scientists must analyze the centromeres of several broadly different species, Willard argues. “Only then can we sit down and ask, What is this telling us?”

>> The Beautiful Weed

At the University of Chicago, Daphne Preuss is doing just that—in a beautiful weed called *Arabidopsis*, which was fully sequenced in 2000, four years after the sequencing of the *Saccharomyces cerevisiae* yeast genome.

Preuss had trained as a yeast geneticist, so when she started studying this plant, which had just been selected as a model organism for biologists, she felt frustrated at not being able to use some of the techniques that were developed for yeast. “We didn't have any of the nice tools I was used to,” she recalls. “I kept complaining and complaining, until finally Ron [Ronald W. Davis of Stanford University, in whose lab she worked as a postdoc] asked me, ‘Why don't you figure out what you can do to make plants better models?’ That struck me as good advice.”

She got her chance in 1993 while looking for plant mutations that affect the shape of pollen. Peering through her microscope, she saw a mutant shape that she found astounding. Ordinarily, *Arabidopsis* plants present a problem for geneticists who wish to track how a cell's chromosomes sort themselves out during meiosis (the cell's division into four eggs or four sperm). A normal plant's chromosomes sort into four separate daughter cells that join a large population of unrelated cells, making such analysis impossible. In the mutant plant, by contrast, the four sex cells were stuck together, as in yeast. This meant that Preuss would be able to track the chromosomes' inheritance and map their centromere region using some of the techniques

that had been used about a decade earlier to locate the “magic sequence” in yeast. “I never had a moment like that—I knew right away it can't get better than that!” Preuss says. “I knew it would fundamentally change everything you could do in plants.”

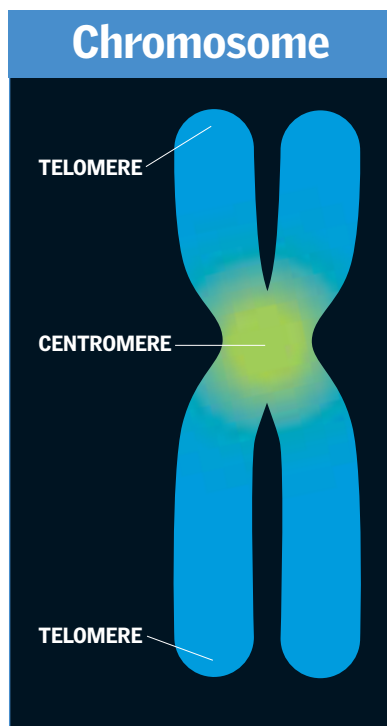
After this epiphany, Preuss temporarily dropped her project on pollination and started to work on *Arabidopsis* centromeres. She took her cues from the research that had been done on yeast a decade earlier. She first crossed plants that contained several DNA variations, which could serve as markers, and then she examined how these markers were partitioned during meiosis. “This involved not only mapping, but knowing the DNA; it involved many people in the Genome Project,” Preuss says. “More of the centromere has [now] been sequenced in *Arabidopsis* than in any other multicellular organism.”

She has also started making an artificial chromosome for her weed and founded a company called Chromatin to do the same in crops. “If artificial chromosomes can become a simple way of delivering genes into plants,” she says, “there could be very important implications for human health.”

As Preuss notes, “The leading cause of death in the world is hunger.

About 30 percent of kids under age five are in danger of starving to death. More than half of the world's soil is bad for cultivation—too salty or too dry or too wet or permafrost, or has chemical or nutrient imbalances. Yet, the genome projects have been very successful in identifying thousands of genes that could have an impact on food production.” These genes could make it possible to grow crops in poor soil, she suggests. Some genes would make plants resistant to insects or fungi; others would make them tolerant to drought. All would enable plants to produce better yields.

“Breeding has been the traditional way to solve such problems, but it is very slow,” she points out. For example, breeders have been searching for years for salt-tolerant plant types, with no success. Genetic modification works more rapidly.



“However, it takes about seven years for one single gene insert to be approved by the FDA, USDA, and EPA for release to farmers, and much longer if you insert several genes separately and then breed them together. Suppose you wanted to change, say, five things—with current technologies, it would take decades!

“The only solution would be to take multiple genes and put them in all at once,” says Preuss. “But for that, you need artificial chromosomes. In our lab, we’ve managed to put four genes into one vector, including a color marker so you can track the chromosome’s presence.”

Preuss is also happy to see that therapeutic drugs are increasingly being grown in plants. “Again, you may need to put more than one gene into the plant to make it practical,” she says. She hopes that artificial chromosomes will enable scientists to do this routinely, given that “plants are such cheap production factories.”

>> The Role of Sexual Competition

When Preuss compared DNA from *Arabidopsis* centromeres to DNA samples from the centromeres of four related plants, she was amazed to see how different their sequences were. Steven Henikoff, at the Fred Hutchinson Cancer Research Center, made similar observations and declared that the sequence repeats found in centromeres are “the most rapidly evolving DNA sequences in the genomes of plants and animals,” differing even between closely related species.

“Such rapid change is paradoxical,” Henikoff and his colleagues wrote in a recent review in *Science*. Since the function of centromeres is universally conserved, they asked, why haven’t centromeres evolved an optimal sequence of DNA? The explanation they suggest is based on the Darwinian competition that occurs during female meiosis, when the centromeres inherited from both parents are duplicated and then go their own way. Only one of the four products of this meiosis will end up in the egg’s nucleus and have a chance at reproduction, the authors note. The other three will become polar bodies and die. The type of centromere that helps a chromosome win this contest (by placing the chromosome in a favorable orientation for the race to reach the egg’s nucleus) will tend to survive and become predominant.

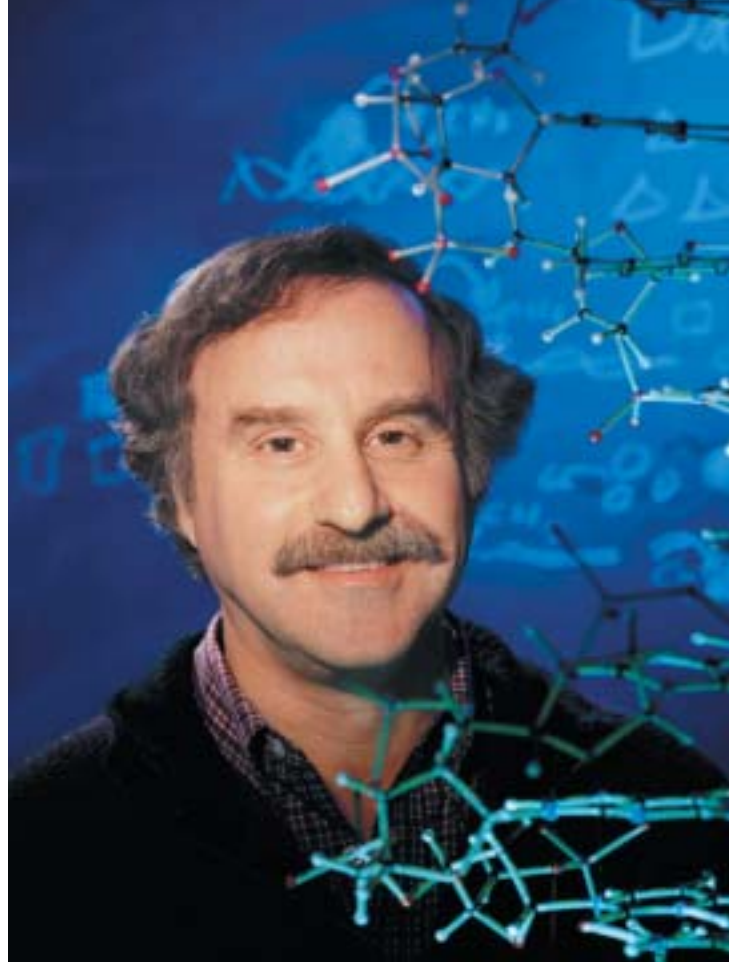
Any “imbalance” between different types of centromeres in an organism might produce problems during male meiosis, however, leading to sterility or other defects. It would need to be counteracted. In *Drosophila*, this job might be performed by the Cid protein, a histone that packages the centromere. Although most histones have a stable, unchanging DNA sequence, this particular histone needs to evolve rapidly in order to attach itself to a changing centromeric sequence and restore the centromere balance.

The result is a kind of “coevolution” between the centromeres and Cid. Henikoff and his colleague Harmit S. Malik call it “an irreversible process of centromere divergence,” in which both the centromere and the protein differ from their ancestors. This theory would explain why centromeric DNA is not conserved across species. It would also solve the long-enduring mystery of why hybrids between species are generally infertile and thus, says Henikoff, “account for the origin of species.”

>> If We Find the Code...

Willard is excited by this theory for an additional reason: He thinks the coevolution of centromere and protein “may be a key part of the code” he and others have been seeking.

Chromosome segregation is about 99 to 99.9 percent effective in



Centromeres are “the most rapidly evolving DNA sequences in the genomes of plants and animals,” differing even between closely related species.

—STEVEN HENIKOFF,

Fred Hutchinson Cancer Research Center

every cell division, he says. “That sounds pretty good, but actually it’s not very good at all. We have about 100 trillion cells in our bodies, so a system that works 99.5 percent of the time is not good enough.

“If you could improve the segregation of human chromosomes by just 10 percent, it would have an enormous effect on the incidence of birth defects and some of the early stages of cancer—the chromosomal changes that underlie tumor growth.”

He suggests focusing on the proteins that interact with the centromeres. “Could we modify these proteins to make the centromeres work better?” he asks. “There could be a pharmaceutical intervention that reduces the frequency of chromosome loss,” which would be very important for pregnant women older than 35 or 40 years, because that’s when the rate of chromosome failure goes up, increasing the risk of having a child with Down syndrome.

“If we find the code,” he says, “we will be able to address the centromeric errors that underlie many birth defects, as well as the chromosomal changes in cancer. And we will have a better chance at intervening.” **H**