

First the human DNA fragments are each spliced into a DNA molecule of a bacterial virus—a bacteriophage, or phage for short. Then these recombinant DNA molecules are each wrapped up in the bodily trappings of the phage, and the phages are allowed to infect some unfortunate bacteria. The bacteria then proceed to work themselves to death making many duplicate viruses, each containing a human DNA fragment in a viral DNA molecule. We gather up the plaques, or colonies of viruses, suspend them in some liquid, and ship them off. So a library actually consists of human DNA fragments encased in bacteriophages.

**HILDEBRAND:** The phage bodies make convenient, protective packages for the fragments. And more important, if a user wants to work one at a time with members of the set of restriction fragments in a library—and that's usually the case—all he needs to do is infect some more bacteria with the phages in his library and pick off one plaque at a time.

**SCIENCE:** *How is such a library used?*

**HILDEBRAND:** It depends on what the user has in mind. In some cases the fragments may be investigated directly—their base sequences determined, for example—and the information pieced together to give some picture of the DNA molecule as a whole. You see, DNA molecules are so huge that details of their organization and functioning can, with present technology, be learned only by studying them in pieces, and preferably not pieces generated by random breaks in the molecules but pieces cut at the same points along each molecule. That's what restriction enzymes do, and it's difficult to overstate how absolutely essential they are to molecular genetics.

Getting back to the question, in most applications a library is used as a source of probes. A probe is a radioactively or fluorescently labeled DNA fragment that is used to detect and locate, in a sample of DNA being investigated, portions of DNA with a complementary base sequence. Of course, any fragment in a library can be labeled and used as a probe, but it may not be detecting anything very exciting. The challenge is to

## What Is a Chromosome?

**T**he hereditary information in each cell of an organism is awesome in extent, complex in organization, and beautiful in the dynamics of its expression. Chromosomes are the packages housing this information. Most of us think of chromosomes as the X-shaped objects viewed through a microscope in a high-school biology course. This familiar manifestation of chromosomes exists, however, only during a certain brief stage in the life of a cell. Since chromosomes—in particular, human chromosomes—are the essential raw material for the Gene Library Project, a brief review of what is now known about these carriers of heredity is worthwhile.

Located in the nucleus, each chromosome includes, as a single molecule, some fraction of the cell's complement of DNA and hence of the cell's genes. The number of chromosomes in a normal cell depends on the species of origin of the cell. Although all organisms of a species have the same chromosome number, that number is not unique to the species. For example, the evening bat (*Nycticeius humeralis*), the red squirrel (*Tamiasciurus hudsonicus streatoris*), and the human all have the same chromosome number. The dependence of chromosome number on species is not straightforward but exhibits a general downward trend with progression up the evolutionary ladder.

In higher organisms another variation in

chromosome number occurs: all gametes (cells that participate in sexual reproduction, such as egg and sperm cells) have the same "haploid" number of chromosomes, and all other cells are "diploid," with exactly twice that number. Human gametes, for example, possess twenty-three chromosomes, and all other human cells possess forty-six.

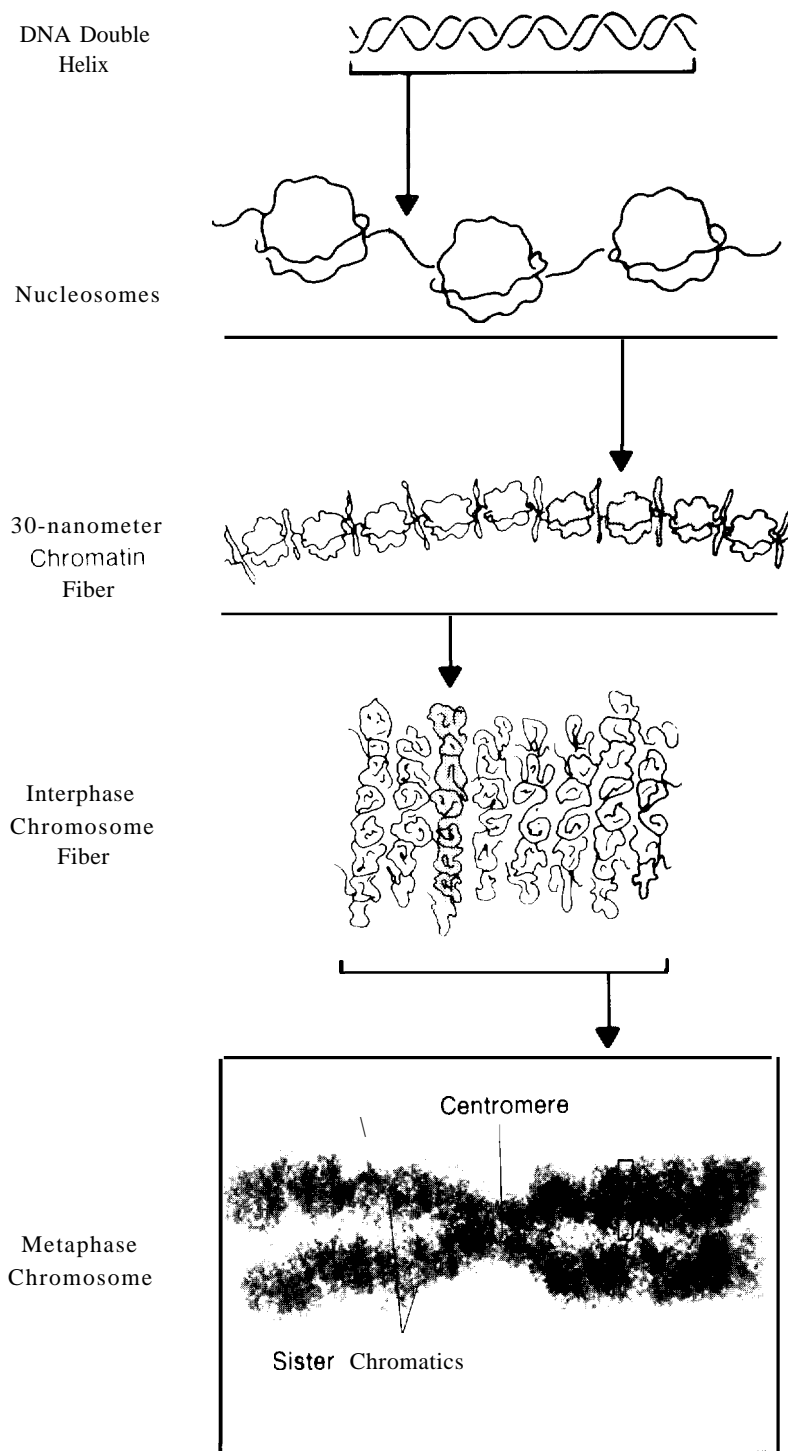
How is the chromosome number of a species maintained constant from one cellular and organismal generation to another? In the case of humans, a new organism begins with the formation of an ancestral diploid cell by fusion of two haploid gametes: an egg carrying an X sex chromosome and a sperm carrying an X or a Y sex chromosome. (The new organism is a female (male) if an X-bearing (Y-bearing) sperm participated in the fusion.) Other cells are then formed by successive waves of mitosis, a type of cell division resulting in two diploid cells, each containing duplicates of the chromosomes of both the maternal and the paternal gametes. (The maternal and paternal versions of each chromosome are referred to jointly as a homologous pair.) The organism grows and develops as mitosis continues and certain of the cells become differentiated in function. At some stage or stages in the organism's life, special diploid cells (germ cells) undergo meiosis, a type of cell division resulting in formation of four haploid gametes. Each chromosome in these gametes is a duplicate of either the

maternal or the paternal chromosome or, more usual, is some combination of the two arrived at by exchange of hereditary material between homologous chromosomes during meiosis. The cycle begins again with fusion of one such gamete with another from an organism of the opposite sex.

A chromosome includes, in addition to a single DNA molecule, a mass of special proteins (histones) roughly equal to that of the DNA molecule (some 15 percent of the total chromosomal mass), small amounts of many other DNA-binding proteins, and RNA. The histones make the DNA molecule more flexible, by neutralizing the negative charges of its phosphate groups, and thus play an essential role in packaging the very long DNA molecules (which, for humans, average 5 centimeters in length) within the confines of the cell nucleus (which is between  $5 \times 10^{-4}$  and  $10^{-3}$  centimeter in diameter). Many details of this remarkable engineering feat are as yet unknown, but several levels of packaging are involved (Fig. 1).

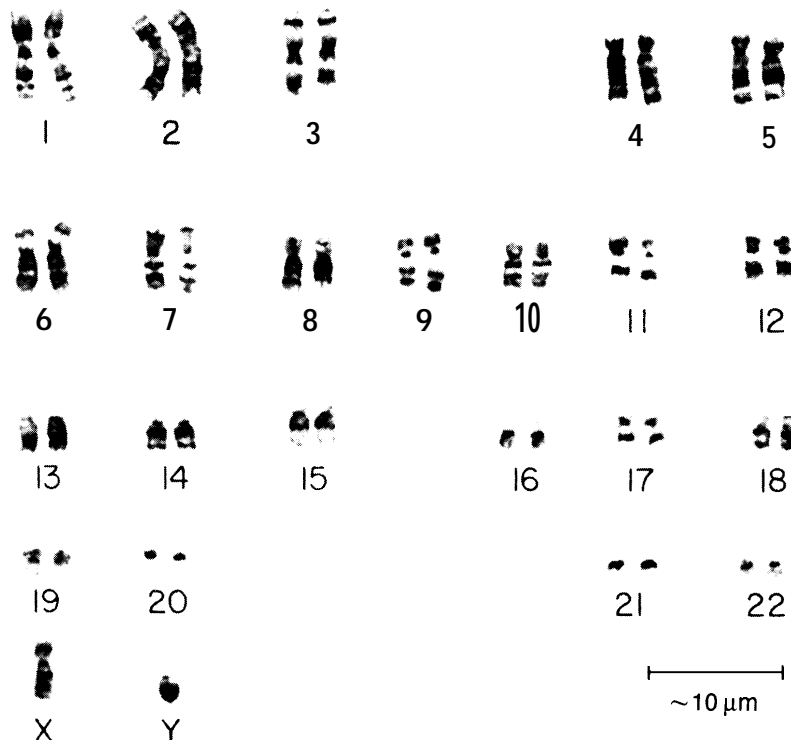
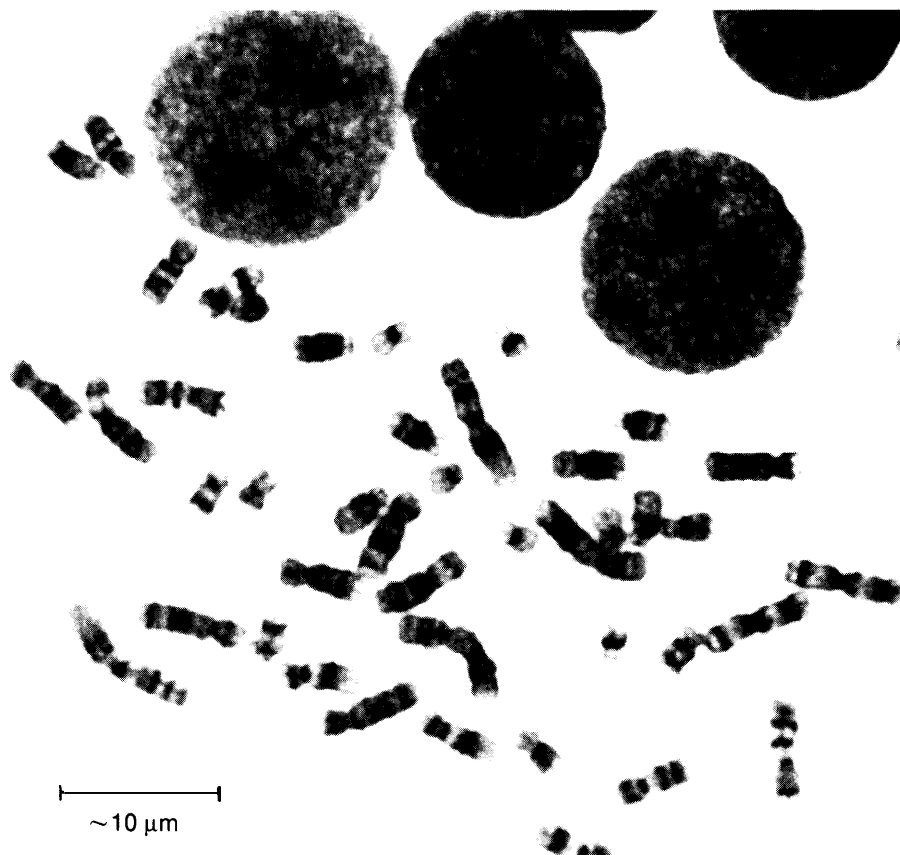
The packaging begins with the formation of "nucleosomes," each consisting of a length of linking DNA and a length of DNA wound around a core of histones. Next, some regular packing of the nucleosomes, also involving histones, leads to the structure known as (the 30-nanometer chromatin fiber (the double helical strand of DNA itself is about 2 nanometers in diameter). This fiber in turn is organized into a sequence of "looped domains," each of which is thought to be activated as a unit during gene expression. The sequence of looped domains is further condensed in some fashion to form the "interphase" chromosome, portions of which decondense and recondense as various genes are activated and deactivated during biosynthesis by the cell.

The ultimate in DNA packaging occurs as a germ cell prepares for meiosis or as a somatic cell (any cell other than a germ cell) prepares for mitosis. Then each interphase chromosome is duplicated, and the two identical "sister chromatids," which are joined at some point along their lengths (the "centromere"), form a longitudinally sym-



**Fig. 1.** Many details of the packaging of a DNA molecule and various proteins into the entity known as a chromosome are yet to be determined. These schematic illustrations are merely suggestive of the mechanisms and configurations that may be involved.

**Fig. 2. Photomicrograph of metaphase chromosomes from human leukocytes (white blood cells). The cells were cultured for 48 hours, blocked at metaphase by exposure to the drug Colcemid for 2 hours, swollen in a hypotonic solution, and treated with a fixative. The chromosomes were released from the metaphase cells simply by dropping the swollen cells onto microscope slides, which breaks the cellular membranes. (The dark, round objects are nuclei released from interphase cells.) The chromosomes were then partially digested with trypsin and treated with Giemsa stain. Since trypsin preferentially digests chromosomal proteins bound to DNA rich in guanine-cytosine pairs, those portions of a chromosome containing such DNA are less readily stained. The result is a pattern of heavily stained (dark) and lightly stained bands along the length of each chromosome. The patterns of these bands permit unambiguous identification of chromosomes.**



**Fig. 3. A display of the metaphase chromosomes of an organism, banded as in Fig. 2 and arranged in order of decreasing size, is known as the karyotype of the organism. Shown here is the karyotype of a human male. By convention the pair of sex chromosomes is displayed separately from the homologous pairs of autosomes. Chromosomes are designated by a size-related number and, historically, by a morphology-related group: group A includes numbers 1 through 3; group B, numbers 4 and 5; group C, numbers 6 through 12; group D, numbers 13 through 15; group E, numbers 16 through 18; group F, numbers 19 and 20; group G, numbers 21 and 22. The similar morphology of the members of a group made them extremely difficult to distinguish before discovery in 1970 of the banding method described in Fig. 2.**

## WHAT IS A CHROMOSOME?

metric double structure with the familiar X shape. These "metaphase," or "mitotic," chromosomes are easily visible through an optical microscope (Fig. 2) and were first observed by German cell biologists in the early 1880s.

The **metaphase** chromosomes of a given species can be distinguished on the basis of a number of properties, such as overall length, DNA content, or location of the centromere. (Except in the case of an X-Y pair, the differences between the two members of a homologous pair of chromosomes are much less pronounced than the differences among homologous pairs.) The most definitive identifying property is the pattern of bands produced on **metaphase** chromosomes by selective digestion of the chromosomal proteins and subsequent staining. Figure 3 displays such banding patterns for the human chromosomes.

Which genes are packaged in which chromosomes, and exactly where, is known only for some 700 of the estimated 100,000 human genes. Examples of those for which this information is known include the gene for Huntington's chorea (chromosome 4), for interferon (chromosome 9), for  $\beta$ -globin (chromosome 11), and for hemophilia and red-green colorblindness (the X chromosome).

The marvelous choreographies of mitosis and meiosis normally transmit intact and unchanged the time-tested chromosomes of an organism from one cellular and **organismal** generation to another. But sometimes errors, such as rearrangements of chromosomes or departures from the normal number, do occur. The consequences of errors in meiosis are usually grave, if not fatal, to the organism that inherits them. For example, an abnormal number of chromosomes are found in the cells of about half of the human embryos that are spontaneously aborted, and the presence of three copies of human chromosome 21 results in Down's syndrome. Errors in mitosis also have serious consequences, as evidenced by the association of chromosome rearrangements with various types of cancer. ■

find the interesting probes, ones that reveal something unusual about the DNA, such as an alteration related to a genetic disease or a mutation.

**SCIENCE:** *Have libraries been produced before?*

**CRAM:** Whole-genome libraries, ones containing DNA fragments from all the human chromosomes, are already available, and so are a few chromosome-specific libraries, ones for the chromosomes that are easier to sort. But this project will be the very first to provide chromosome-specific libraries for all twenty-four types of human chromosomes.

**VAN DILLA:** With a chromosome-specific library a user can begin his research by fishing around among a pool of DNA fragments from a particular chromosome rather than the whole sea of human DNA. This is a tremendous advantage, and the libraries should increase the productivity of research in molecular genetics just as separated isotopes increase the productivity of research in nuclear physics,

**DEAVEN:** I have some data to support that expectation, at least as far as one application of the libraries—gene mapping—is concerned. By gene mapping is meant localizing a gene on a particular region of a particular chromosome. In 1958, when molecular genetics was still in its infancy, only about four hundred human genes had been identified. These had been classified as dominant or recessive and autosomal or X-linked but, except for the X-linked genes, could not be mapped. That had to await, first of all, the discovery some ten years later of chromosome banding, a technique for unambiguously distinguishing all twenty-four types of human chromosomes. This discovery made it possible to map a limited number of genes by associating chromosome abnormalities and genetic diseases. At about the same time human-rodent hybrid cell lines became available, and information about the proteins synthesized by various of these cell lines, each carrying only a few human chromosomes, made it possible to map more genes. Still, by 1973 only 64 genes had been assigned to particular autosomes and 15510

the X chromosome. By 1981 the total number of mapped genes had increased to only about four hundred and fifty, but in that year molecular genetics provided a powerful new tool for mapping. It was discovered that what are known as restriction fragment length polymorphisms could pinpoint the locations of genes about which practically nothing is known except the observable evidence of their expression, such as the symptoms of diseases.

A restriction enzyme cuts a DNA molecule into fragments of various lengths by breaking bonds at every occurrence of a cleavage site, a certain sequence of bases. Mutations can generate new cleavage sites and destroy existing ones, and these changes cause differences in the sets of restriction fragments from different individuals. For example, in the fragments produced by the restriction enzyme *HpaI* from the DNA of American Blacks, the  $\beta$ -globin gene usually appears on a 7.6-kilobase fragment but sometimes on a 7.0- or 13.0-kilobase fragment. Differences like these are called restriction fragment length polymorphisms, or RFLPs. An RFLP can be detected by using electrophoresis to separate, according to length, the restriction fragments from, say, two individuals and seeing whether a probe that includes the RFLP sticks to different places in the two lineups.

An RFLP within a gene is obviously inherited along with the gene, but so also is an RFLP so close to a gene that it is not separated from it during meiotic recombination. So the location of such an RFLP on a chromosome is a very close approximation to that of the gene. This method of mapping, say, for a gene that causes some disease, involves testing many probes to find one that reveals an RFLP unique to people with and without the disease—that's the hard part—and then mapping the probe on the chromosome—that's the easy part.

Since 1983 about three hundred RFLP probes for genes have been identified and mapped per year. Even that rate, though, is painfully slow considering how many more genes remain to be mapped. But our chromo-