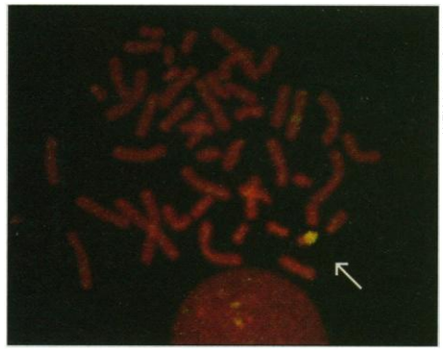
Two Chromosomes Down, 22 to Go

Researchers have produced detailed physical maps of human chromosomes Y and 21, providing a boost both to the Human Genome Project and to efforts to locate disease genes

As the Human Genome Project officially turns 2 years old this month, it has hit its stride even sooner than its most ardent enthusiasts had predicted. Data are pouring out of the genome centers, new technologies are coming on line, and, perhaps most notably, the first two high-resolution maps of human chromosomes are now complete.

David Page's group at the Whitehead Institute reports on pages 52 and 60 that they've cloned and mapped the mysterious Y or male chromosome. And Daniel Cohen at the Centre d'Etude du Polymorphisme Humain (CEPH) and its companion lab Généthon in Paris, reports in the 1 October *Nature* that his group has mapped chromosome 21, long the focus of intense study because it houses genes involved in Down's syndrome, Alzheimer's, and other neurological diseases. Each map is an actual physical representation of the chromosome, consisting of chunks of cloned DNA pieced together in the correct order. At a time when the genome project is under increasing budget scrutiny, the maps are proof positive that this big-ticket item "is coming through with the goods," says David Cox, a genome mapper at the University of California, San Francisco.

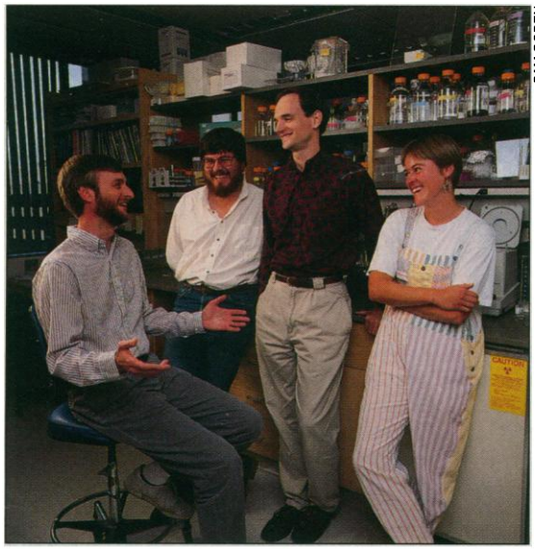
And the goods are numerous. Already, the maps are an immense boon to researchers hunting genes on either chromosome. For these two, at least, it means the agonizing chromosome "walks" and "jumps" previously needed to close in on a long-sought gene—which can cost $20 million or $30 million and consume several years—are over. Instead, once genetic linkage studies have narrowed the hunt to a few million base pair region of the chromosome, investigators can simply go to the freezer and pull out that particular piece of DNA. They must still find a way to ferret out the gene itself, but with the map in hand, says David Botstein of Stanford, "It is the difference between finding a book in a library with a catalog or without."



WEN-LIN KUO, UCSF

Tiny target. The Y chromosome (arrow) is the smallest of the human chromosomes.

In terms of the genome project, these maps mean that it may actually meet or even exceed its ambitious goals to complete physical maps of all the human chromosomes in 5 years. "This is the kind of milestone the genome project was waiting for," says geneticist Francis Collins of the University of Michigan, who is rumored to be the leading candidate to replace James Watson as head of the NIH genome effort. "I think everyone will breathe a sigh of relief. Physical mapping is moving ahead faster than anyone predicted."



SAM OGDEN

Y mappers. From left to right, they are David Page, Simon Foote, Douglas Vollrath, and Adrienne Hilton.

Admittedly, these are among the smallest chromosomes—together they account for about 2% of the human genome. And the maps are not perfect, or even complete. (Both Page and Cohen skipped the portions of the chromosomes composed of repeated sequences, which are hard to map and biologically far less interesting.) But even so, Page and Cohen's achievements are particularly sweet because they vindicate a novel mapping strategy that no one was sure would work even a few years ago. In it, maps are assembled from giant clones, known as YACs, with the aid of a new type of marker that can be detected by the polymerase chain reaction (PCR).

But while the Page and Cohen groups both used an almost identical mapping approach, their efforts were prompted by very different goals. Page was driven by his longstanding fascination with the Y chromosome, perhaps the least understood of all the chromosomes. Cohen, on the other hand, set out to push the technology as far and fast as he could, and he mapped 21 simply to test the new megaclones his group had constructed. Both, too, draw very different lessons from their experiences. To Page, like most of the U.S. genome community, these maps chart a clear course on how to proceed. To Cohen, however, the arduousness of the task convinced him that there must be a cheaper and faster way to go—and he has already set out to find it.

Probing the mysterious Y

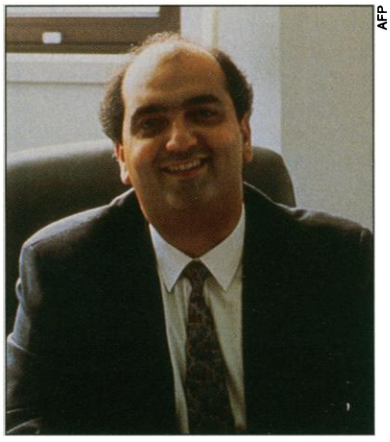
Page has spent the past 10 years trying to sort out the peculiar biology of the Y chromosome, which he calls a "mysterious jungle." Until recently many geneticists thought it consisted entirely of "junk" DNA. Now it is clear that the Y does contain genes, but how many and what they might do is anyone's guess. The Y is so poorly understood because it is not amenable to one of the most efficient tools for localizing genes, genetic linkage mapping. Such maps depend on analyzing chromosomes when they divide and recombine during the formation of egg and sperm—but most of the Y does not undergo recombination. The upshot, says Page, is that "especially for the Y, we have to work from the DNA up to understand function. The biology will never be understood without a physical map."

Building such physical maps is much like solving a jigsaw puzzle. First the DNA is chopped into pieces with restriction enzymes and then copied, or cloned. Then the pieces—the DNA clones—are assembled by looking for overlapping patterns. Earlier mapping efforts, on yeast and the roundworm *Caenorhabditis elegans*, for instance, were hamstrung by limitations on the size of the DNA segments that could be cloned—no more than 40 kilobases at best. That meant that tens of thousands of clones might have to be analyzed to map an average chromosome. Obvi-

ously, the bigger the pieces, the easier the task, and that is where YACs come in.

In 1987, Maynard Olson, then at Washington University, and his graduate student David Burke modified yeast chromosomes to create YACs (for yeast artificial chromosomes) into which stretches of DNA 250 kilobases long could be inserted. These inserts would be copied every time the yeast cells divided. This achievement changed the prospects for large-scale physical mapping almost overnight. Still, a number of questions remained. Would human DNA be stable in YACs, for instance? And could overlapping pieces be linked together into one giant set of overlapping clones—known as a contig—spanning millions of bases? To Page it was worth a gamble, and in 1990 he and Simon Foote, Douglas Vollrath, and Adrienne Hilton plunged in, focusing on the so-called euchromatic region of the Y that contains the genes and not the heterochromatic region full of repetitive DNA.

Their first task was to see if they could assemble even bigger YACs—an arduous task that consumed much of the 18-month project, says Page. To give themselves a leg up, the Whitehead group started with DNA from a male with three extra copies of his Y chromosome, eventually achieving clones 650 kilobases long, as Cohen's group had recently done.



AFP

Bold plan. Daniel Cohen wants to map the whole genome.

To turn this collection of jumbled clones into an ordered array, the Whitehead group used the relatively untested technique of STS (for sequence-tagged sites) content mapping. They first generated nearly 200 STS markers, which are short unique stretches of DNA that can be detected by PCR and mapped to a particular point on the chromosome. Next they established the markers' relative order on a deletion map, a cruder kind of chromosome map that Page's group has also been working on. Then the group screened the YAC clones they had made, using the STS markers both to identify the Y clones and simultaneously put them in the correct order. The underlying premise is that if two clones share one or more STS markers, they must overlap.

It worked. "It is all one process," says Page. "It is fairly amazing to see the chromosome coalesce." And much to their surprise, screening turned out to be largely "a turn of the crank operation," he says—time-consuming but fairly straightforward. The end result is a map that covers 28 million bases with overlapping clones—and no apparent gaps. The STS markers are spaced roughly every 220 kilobases, slightly shy of the 100-kilobase resolution the genome project called for in its 5-year plan.

With this new map, says Page, "we can finally think about finding all the genes on the Y." And the map will also give added impetus to the quest for what Page hesitantly dubs the "Y chromosome Adam," a reference to the late Allan Wilson's "mitochondrial Eve." As Page explains, geneticists can use Y DNA, which is inherited only from the father, in much the same way that mitochondrial DNA, inherited only from the mother, is used to reconstruct human genealogies.

Thinking big

At about the time that the Page group was embarking on the Y map, Cohen, Ilia Chumakov, and colleagues launched a huge effort to develop giant YACs for physical mapping. In early 1992, Cohen astounded the genome community when he achieved megaYACs, with an average insert size of 1.2 to 1.4 million bases. Cohen is the first to admit that they employed no stunning new technology or strategy. The trick, he says, was "just manpower and morale," with five people working on making YACs full time. And to this day he has no idea how they pulled it off. The process is "highly unreproducible," he says. "We never know when it will work, even for us."

With the megaYACs in hand, Cohen wanted to test them to see if any of the inherent problems in YACs, such as rearrangement or instability, would impede mapping. The chromosome 21 map is the result, created by "absolutely the same" methods as the Page group used, says Cohen. It spans the entire long arm, about 43 million bases, from the centromere to the very tip of the chromosome, with STS markers spaced every 200 kilobases. Cohen's group did not bother mapping the short arm, which consists of repetitive DNA sequences. As with Page's Y map, there are no visible gaps.

Page and Cohen are quick to point out that there are two important qualifiers for both maps. Although no gaps are obvious in either, some of the DNA is probably missing—in other words, stretches of DNA either were not cloned in the first place or, more likely, were subsequently deleted from the YACs. In addition, both groups encountered certain STS markers for which they could not establish a definitive order. In Cohen's map, just seven of the 198 STS markers proved problematic. "We know exactly [the region in which] they fall on the map, but we don't know the order."

To genome researchers like Collins and Olson, these are just minor nits. Both groups managed for the first time to map tens of millions of base pairs this way, in a continuous set of clones. And if you do that, asserts Olson, "you can map anything." What's more, both groups pulled it off without fancy bells and whistles or mammoth teams: The Y map was done by three people over 18 months; the 21 map by four people over a year. The secret of Page and Cohen's success, say their colleagues, was their singlemindedness. To Cox, the take-home message is: "If you just roll up your sleeves you can do it."

Ironically, to Cohen, Cox's comment underscores what's wrong with the U.S. goal of mapping each chromosome this way, with even more closely spaced STS markers: It is far too tedious and expensive. Based on his experience, Cohen is convinced that is the wrong way to go—or, as he says more diplomatically, "it is not the approach for us. We believe in the whole genome approach," made possible by his megaYACs.

Cohen argues that generating and sorting clones and markers for each chromosome is far too inefficient. It makes more sense, he says, to tackle all the chromosomes at once. Cohen's plan is to construct clones and markers from the entire genome, taking advantage of economies of scale, and then use a different "fingerprinting" technique to look for overlaps and build up contigs without worrying which chromosome they fall on. The idea is to assemble ever larger contigs until they are big enough to map back to the chromosomes quickly.

In keeping with his bold style, Cohen is already moving ahead. As he and colleagues reported in *Cell* last week, they have already cloned half the genome this way, in 1000 pieces or contigs. He predicts the entire map will be done within a year. True, this will be a low-resolution map, without the STS markers that are so useful in tracking down genes and comparing data. But in Cohen's book, "it is better to use cheap methods to get the full map and then start sequencing as much as possible."

Clearly, with both megaYACs and chromosome 21 under his belt, Cohen is a force to be reckoned with. In part prompted by his success with megaYACs, a number of U.S. groups, including Page's, are now leaning toward the whole genome approach, though few are ready to give up STSs. "He certainly makes a strong case for the elegance and speed of that approach," says Collins. "We should know in a year if it works." Whatever the outcome, few would have predicted when the genome project began that in 2 years they would be contemplating mapping the entire genome in one fell swoop.

–Leslie Roberts