

JUNE 19, 2003

Finished Y-Chromosome Sequence Reveals a Genomic "Crystal Palace"

A team of 40 researchers has finished sequencing the Y chromosome, the male sex chromosome once belittled as “the Rodney Dangerfield of the human genome” because researchers believed it contained no genes of interest.

The Y may gain a measure of respect now that researchers have discovered that it is actually a genomic “crystal palace,” containing genes that impact male fertility, vast stretches of mirror-image DNA, and an assortment of functional and vestigial genes.

"I have been told for years that the Y chromosome was full of junky repeats, a genetic wasteland."

— **David C. Page**

Most significantly, the new studies have unearthed a startling mechanism that the Y chromosome uses to maintain its functionality. It appears that the Y protects its genetic integrity by swapping multiple copies of the same gene within its own structure.

“I have been told for years that the Y chromosome was full of junky repeats, a genetic wasteland,” said senior author David C. Page, a Howard Hughes Medical Institute investigator at MIT's Whitehead Institute for Biomedical Research. “People would ask why we would consider wasting our time mapping and sequencing it. But, in fact, what we see is that it's a crystal palace.”

The researchers published their findings in two articles in the June 19, 2003, issue of the journal *Nature*. Page collaborated with colleagues from the Whitehead Institute, Washington University School of Medicine and the Academic Medical Center in Amsterdam.

According to Page, the team's detailed genomic sequence of the human Y chromosome will contribute to a better understanding of male infertility, as well as certain sex-linked genetic disorders in women. He also speculated that their findings could lead to genomic explanations for differences in disease

susceptibility between men and women.

Sex chromosomes in animals and humans include the X chromosome and the much smaller Y chromosome. Females have a pair of X chromosomes and males have both an X and Y chromosome.

Although the sequences of several organisms with Y chromosomes—including fruit flies and mice—have been declared “complete,” said Page, those genome sequences did not include the Y chromosome. In the case of the human Y chromosome—and likely those of other animals—standard sequencing techniques have been confounded by the long stretches of nearly identical DNA that lack the landmarks needed to guide the assembly of sequences from smaller segments, said Page.

The Y is unique among chromosomes in that it bears long stretches of mirror-image, or “palindromic,” DNA sequences that nearly thwarted sequencing efforts. “Nobody has seen palindromes of this scale and degree of precision anywhere in the genome,” said Page. “Before we began this project, when asked why it would be so hard to map and sequence the Y chromosome, I used to say it feels like a hall of mirrors. I had no idea how accurate that analogy was, because the Y literally *is* a hall of mirrors. Trying to sequence the Y chromosome is, indeed, like entering a hall of mirrors, spinning around, and after coming outside, being asked to draw the hall's floor plan. You're disoriented.”

To overcome the disorienting duplications in DNA sequence, Page and his colleagues employed an iterative method to get an overall picture of the Y map. They later refined the technique to do more precise sequencing of the individual segments of the Y chromosome. By sequencing the Y chromosome in segments, they were able to detect minute differences among the near-identical palindromes. They then folded that sequence data back into their mapping to improve it.

In order to minimize complications due to normal genetic variability among males, the sequencing was performed on the Y chromosome of one male, whose identity remains anonymous.

Page said much of the credit for the high-precision sequencing goes to the Washington University Genome Sequencing Center, whose researchers achieved an overall accuracy in the Y-chromosome sequence of one error in 100,000 to 1,000,000 DNA base pairs. This feat is even more impressive when one considers that in order to deduce the approximately 24-million-base-pair sequence of the Y chromosome, the team had to sequence well over 50 million base pairs of DNA, according to Page's estimates.

The final sequence reveals that the Y chromosome is a mosaic of two kinds of genomic sequences: euchromatic sequences that represent active genes, and heterochromatic sequences that are nonfunctional.

The functional euchromatic sequences included three classes, said Page. “These three classes really shout out messages about the evolution of the Y chromosome and sex chromosomes in general, and about the function of the Y chromosome today,” he said.

The three classes are called “X-degenerate,” “X-transposed” and “ampliconic” sequences. X-degenerate sequences are relics from an ancient time when the X and Y chromosomes first evolved from an ordinary, or autosomal, chromosome. Genes within these sequences—which resemble genes on the X chromosome—show evidence of steady decay due to mutation, and many are non-functional. “We can see evidence that, whereas the gene on the X is a functional working copy, in many cases the corresponding gene on the Y is a rotted-out hulk that no longer does any business,” said Page. “And that provides us a glimpse into one aspect of the Y chromosome as a rotted-out X chromosome.”

The X-transposed sequences are genes that were swapped as a group from the X chromosome roughly three to four million years ago, after the ancestors of humans and chimpanzees diverged into separate lines. There are few functional genes in this region, said Page.

Finally, the ampliconic sequences are those that exist within multiple, repeated palindromic segments. “The ampliconic genes are the big surprise,” said Page. “While the X-degenerate genes tend to be expressed throughout the body and in many different tissues and cell types, the genes in the ampliconic sequences are very restricted to the testis in their expression. And to the extent that we've studied them in detail, it looks like they're actually expressed in only in the spermatogenic cells themselves.” Thus, said Page, these genes are likely to play an extremely important role in generating sperm. This role has been confirmed by earlier work showing that mutations in the Y chromosome are the most common known genetic causes of male infertility.

Perhaps the most fundamental insight arising from the sequencing of the Y chromosome, said Page, is how ampliconic genes avoid degradation due to mutation. Unlike the two X chromosomes in females, the Y chromosome does not have a partner with which to swap genes during cell division in order to replace genes that have suffered deleterious mutations, said Page.

“This was the theoretical underpinning for the traditional notion that the Y was a genetic wasteland—the Rodney Dangerfield of the genome,” said Page. “But we believe we have found that many of the genes on the Y, and virtually all the ampliconic genes, occur in pairs. And so, pairs of genes on the Y can swap, not with genes on another chromosome, but with a partner on the corresponding identical palindrome. This Y-Y gene conversion is, I think, the most important finding of our work.” However, he added, that same internal recombination underlies the chromosomal aberrations that lead to male infertility.

To confirm that the ampliconic genes on the Y chromosome palindromes have been recombining over time, Page and his colleagues also performed a

comparative analysis of Y chromosomes of humans and chimpanzee sequences in those regions. As reported in the second *Nature* paper, that comparison, indeed, revealed that such recombination exists in both species.

“The sex chromosomes represent a grand experiment of nature,” added Page. “And in our work, every few years we've caught a glimpse of some unexpected aspect of this experiment. And of all these aspects, this Y-Y gene conversion is one of the wildest.”

Page emphasized that the scientific and clinical implications of the sequencing of the Y chromosome are profound. For example, comparative sequencing of the Y chromosome among various human populations will reveal much about its variation and functions.

More broadly, he said, “While the sequencing of the human genome has been extraordinarily valuable, I think our work illustrates that those hardest parts of the genome that remain to be sequenced might house particular gems worth finding.” And further digging in especially frustrating heterochromatic regions of DNA, which also contain massively duplicated blocks of genetic material, might also yield new genomic insights.

Clinicians are already using data from the Y-chromosome sequencing to understand the genetic origins of male infertility, said Page. That genomic data will aid in understanding Turner syndrome, one of the most common chromosomal disorders in females. The disorder arises from the lack of one sex chromosome, and the absent gene might well be an X-degenerate gene or its counterpart on the X chromosome, said Page.

On a more speculative note, Page said that genes on the Y chromosome might play a role in influencing gender-specific differences in disease susceptibility. Evidence has developed that the Y chromosome plays a role in gonadal sex determination, skeletal growth, germ-cell tumorigenesis and graft rejection, he said.

“We know that there are many diseases for which men or women are at higher risk,” said Page. “It has been conventionally assumed that these differences in disease susceptibility reflect the action of sex hormones, and not the action of the sex chromosomes directly.” But that assumption was set in place when some people thought the Y didn't have any genes on it, said Page.

At one time, researchers believed that during development of females, all genes on one X chromosome were inactivated, leaving only one full complement of genes on the other X chromosome. And since the Y supposedly harbored no genes other than reproductively related ones, men and women were supposedly genetically equivalent.

“But now we know there are many genes on the X that escape inactivation, so they are present in two copies in females and in one copy in males. So, maybe we should rethink the roles of the second sex chromosomes in these often dramatic differences in disease susceptibility between males and

females.”