February 13, 2001, Tuesday

READING THE BOOK OF LIFE; Genome's Riddle: Few Genes, Much Complexity

By NICHOLAS WADE

The human genome is the most precious body of information imaginable. Yet the biologists who yesterday reported their first analysis of the decoded sequence have found as much perplexity as enlightenment.

The chief puzzle is the apparently meager number of human genes. Textbooks have long estimated 100,000, a number that seemed perfectly appropriate even after the first two animal genomes were deciphered. The laboratory roundworm, sequenced in December 1998, has 19,098 genes and the fruit fly, decoded last March, owns 13,601 genes. But the human gene complement has now turned out to be far closer to genetic patrimony of these two tiny invertebrates than almost anyone had expected.

Dr. J. Craig Venter and colleagues at Celera Genomics report in this week's Science that they have identified 26,588 human genes for sure, with another 12,731 candidate genes. When they first screened the gene families likely to have new members of interest to pharmaceutical companies, "there was almost panic because the genes weren't there," Dr. Venter said.

Celera's rival, the publicly funded consortium of academic centers, has come to a similar conclusion. Its report in this week's Nature pegs the probable number of human genes at 30,000 to 40,000. Because the current gene-finding methods tend to overpredict, each side prefers the lower end of its range, and 30,000 seems to be the new favorite estimate.

The two teams, who discussed their findings in a news conference yesterday in Washington, found other oddities, too. Most of the repetitive DNA sequences in the 75 percent of the genome that is essentially junk ceased to accumulate millions of years ago, but a few of sequences are still active and may do some good. The chromosomes themselves have a rich archaeology. Large blocks of genes seem to have been extensively copied from one human chromosome to another, beckoning genetic archaeologists to figure out the order in which the copying occurred and thus to reconstruct the history of the animal genome.

As the modest number of human genes became apparent, biologists in both teams were forced to think how to account for the greater complexity of people, given that they seem to possess only 50 percent more genes than the roundworm. It is not foolish pride to suppose there is something more to Homo sapiens than Caenorhabditis elegans. The roundworm is a little tube of a creature with a body of 959 cells, of which 302 are neurons in what passes for its brain. Humans have 100 trillion cells in their body, including 100 billion brain cells.

Several explanations are emerging for how to generate extra complexity other than by adding more genes. One is the general idea of combinatorial complexity -- with just a few extra proteins one could make a much larger number of different combinations between them. In a commentary in Science, Dr. Jean-Michel Claverie, of the French National Research Center in Marseille, notes that with a simple combinatorial scheme, a 30,000-gene organism like the human can in principle be made almost infinitely more complicated.

But Dr. Claverie suspects humans are not that much more elaborate than some of their creations. "In fact," he writes, "with 30,000 genes, each directly interacting with four or five others on average, the human genome is not significantly more complex than a modern jet airplane, which contains more than 200,000 unique parts, each of them interacting with three or four others on average."

The two teams' first scanning of the genome suggests two specific ways in which humans have become more complex than worms. One comes from analysis of what are called protein domains. Proteins, the working parts of the cell, are often multipurpose tools, with each role being performed by a different section or domain of the protein.

Many protein domains are very ancient. Comparing the domains of proteins made by the roundworm, the fruit fly and people, the consortium reports that only 7 percent of the protein domains found in people were absent from worm and fly, suggesting that "few new protein domains have been invented in the vertebrate lineage."

But these domains have been mixed and matched in the vertebrate line to create more complex proteins. "The main invention seems to have been cobbling things together to make a multitasked protein," said Dr. Francis S. Collins, director of the genome institute at the National Institutes of Health and leader of the consortium. "Maybe evolution designed most of the basic folds that proteins could

use a long time ago, and the major advances in the last 400 million years have been to figure out how to shuffle those in interesting ways. That gives another reason not to panic," he said, in reference to fears about the impoverished genetic design of humans.

Evolution has devised another ingenious way of increasing complexity, which is to divide a gene into several different segments and use them in different combinations to make different proteins. The protein-coding segments of a gene are known as exons and the DNA in between as introns. The initial transcript of a gene is processed by a delicate piece of cellular machinery known as a spliceosome, which strips out all the introns and joins the exons together. Sometimes, perhaps because of signals from the introns that have yet to be identified, certain exons are skipped, and a different protein is made. The ability to make different proteins from the same gene is known as alternative splicing.

The consortium's biologists say that alternative splicing is more common in human cells than in the fly or worm and that the full set of human proteins could be five times as large as the worm's. Another possible source of extra complexity is that human proteins have sugars and other chemical groups attached to them after synthesis.

There's a different explanation of human complexity, which is simply that the new low-ball figure of human genes derived by Celera and consortium is a gross undercount. Dr. William Haseltine, president of Human Genome Sciences, has long maintained that there are 120,000 or so human genes. Dr. Randy Scott, chief scientific officer of Incyte Genomics, predicted in September 1999 that there were 142,634 human genes. Last week Dr. Scott said he accepted the rationale for the lesser number and now puts the human complement at around 40,000.

Dr. Haseltine, however, remains unshaken in his estimate of 100,000 to 120,000 genes. He said last week that his company had captured and sequenced 90,000 full-length genes, from which all alternative splice forms and other usual sources of confusion have been removed. He has made and tested the proteins from 10,000 of these genes. The consortium and Celera have both arrived at the same low number because both are using the same faulty methods, in his view.

"I believe their gene finding methods are far more imperfect than they own up to," Dr. Haseltine said, noting that 5 of the 10 genes in the AIDS virus were missed at first. "It's my personal conviction that as further studies of chromosomes continue the number of genes will rise until they match the number we project of 100,000 to 120,000."

Dr. Haseltine notes that the gene- finding methods used by the two teams depend in part on looking for genes like those already known, a procedure that may well miss radically different types of genes. His own method, capturing the genes produced by variety of human cell types, is one that Dr. Venter says in his paper is the ultimate method of counting human genes.

Dr. Haseltine is at present in a camp of one. Dr. Venter strongly disagrees, as do members of the consortium. Dr. Eric S. Lander of the Whitehead Institute last week challenged Dr. Haseltine to make public all the genes he had found in a 1 percent region of the genome and let others assess his claim. Dr. Collins said that there was "a terrific way to size up his claims -- let an objective third party look at the data."

"I'd be glad to help arrange that," he said.

Dr. Haseltine said yesterday that he was contemplating the best way to respond and that he was "planning to do so in one form or another, in the open literature."

Turning from genes to chromosomes, one of the most interesting discoveries in this week's papers concerns segmental duplications, or the copying of whole blocks of genes from one chromosome to the other. These block transfers are so extensive that they seem to have been a major evolutionary factor in the genome's present size and architecture. They may arise because of a protective mechanism in which the cell reinserts broken-off fragments of DNA back into the chromosomes.

In Celera's genome article, Dr. Venter presents a table showing how often blocks of similar genes in the same order can be found throughout the genome. Chromosome 19 seems the biggest borrower, or maybe lender, with blocks of genes shared with 16 other chromosomes.

Much the same set of large-scale block transfers seems to have occurred in the mouse genome, Dr. Venter writes, suggesting that the duplications "appear to predate the two species' divergence" about 100 million years ago. He hopes that by sequencing the genomes of many other species he can reconstruct the history of the genome's formation.

Segmental duplication is an important source of innovation because the copied block of genes is free to develop new functions. An idea enshrined in many textbooks is that the whole genome of early animals has twice been duplicated to form the vertebrate lineage. There are several cases in which one gene is found in the roundworm or fly and four very similar genes in vertebrates. (The quadruplicated genes that failed to find a useful role would have been shed from the genome.)

But neither Celera nor the consortium has found any evidence for the alleged quadruplication. If this venerable theory is incorrect, the four-gene families may all arise from segmental duplication.

No one could expect a text as vast and enigmatic as the human genome to yield all its secrets at first glance, and indeed it has not done so. Dr. Venter said that the principal purpose of his paper was to describe the sequence and that he would convene conferences of experts to help further interpret it.

Dr. Lander said the consortium's analysis too was just preliminary. "We tried to write a paper that was not the last word on the genome but sketched all the directions you could go in," he said. "The goal was to launch a thousand ships, not to catalog a thousand genes."

Organizations mentioned in this article:

Related Terms:
Genetics and Heredity; Proteins; Biology and Biochemistry
