The New York Times
nytimes.com

PRINTER-FRIENDLY FORMAT
SPONSORED BY

(500) DAYS OF SUMMER

December 12, 2006

SCIENTIST AT WORK | NICK PATTERSON

A Cryptologist Takes a Crack at Deciphering DNA's Deep Secrets

By INGFEI CHEN

Thirty years ago, Nick Patterson worked in the secret halls of the Government Communications Headquarters, the code-breaking British agency that unscrambles intercepted messages and encrypts clandestine communications. He applied his brain to "the hardest problems the British had," said Dr. Patterson, a mathematician.

Today, at 59, he is tackling perhaps the toughest code of all — the human genome. Five years ago, Dr. Patterson joined the Broad Institute, a joint research center of Harvard and the Massachusetts Institute of Technology. His dexterity with numbers has already helped uncover startling information about ancient human origins.

In a study released in May, scientists at the Broad Institute scanned 20 million "letters" of genetic sequence from each of the human, chimpanzee, gorilla and macaque monkey genomes. Based on DNA differences, the researchers speculated that millions of years after an initial evolutionary split between human ancestors and chimp ancestors, the two lineages might have interbred again before diverging for good.

The controversial theory was built on the strength of rigorous statistical and mathematical modeling calculations on computers running complex algorithms. That is where Dr. Patterson contributed, working with the study's leader, David Reich, who is a population geneticist, and others. Their findings were published in Nature.

Genomics is a third career for Dr. Patterson, who confesses he used to find biology articles in Nature "largely impenetrable." After 20 years in cryptography, he was lured to Wall Street to help build mathematical models for predicting the markets. His professional zigzags have a unifying thread, however: "I'm a data guy," Dr. Patterson said. "What I know about is how to analyze big, complicated data sets."

In 2000, he pondered who had the most interesting, most complex data sets and decided "it had to be the biology people."

Biologists are awash in DNA code. Last year alone, the Broad Institute sequenced nearly 70 billion bases of DNA, or 23 human genomes' worth. Researchers are mining that trove to learn how humans evolved, which mutations cause cancer, and which genes respond to a given drug. Since biology has become an information science, said Eric S. Lander, a mathematician-turned-geneticist who directs the Broad Institute, "the premium now is on being able to interpret the data." That is why quantitative-minded geeks from mathematics, physics and computer science have flocked to biology.

Scientists who write powerful DNA-sifting algorithms are the engine driving the genomics field, said Edward M. Rubin, a geneticist and director of the federal Joint Genome Institute in Walnut Creek, Calif. Like the Broad, the genome institute is packed with computational people, including "a bunch of astrophysicists who somehow wandered in and never left," said Dr. Rubin, originally a physics major himself. Most have never touched a Petri dish.

Dr. Patterson belongs to this new breed of biologist. The shelves of his office in Cambridge, Mass., carry arcane math titles, yet he can converse just as deeply about Buddhism or Thucydides, whose writings he has studied in ancient Greek. He is prone to outbursts of boisterous laughter.

He was born in London in 1947. When he was 2 his Irish parents learned that he had a congenital bone disease that distorted the left side of his skull; his left eye is blind. He became a child chess prodigy who earned top scores on math exams, and later attended Cambridge, completing a math doctorate in finite group theory. In 1969, he won the Irish chess championship.

In 1972, Dr. Patterson began working at the Government Communications Headquarters, where his research remains classified. He absorbed through his mentors the mathematical philosophy of Alan Turing, the genius whose crew at Bletchley Park — the headquarters' predecessor — broke Germany's encryption codes during World War II. The biggest lesson he learned from Dr. Turing's work, he said, was "an attitude of how you look at data and do statistics."

In particular, Dr. Turing was an innovator in Bayesian statistics, which regard probability as dependent upon one's opinion about the odds of something occurring, and which allows for updating that opinion with new data. In the 1970s, cryptographers at the communications headquarters were harnessing this approach, Dr. Patterson said, even while academics considered flexible Bayesian rules heretical.

In 1980, Dr. Patterson moved with his wife and children to Princeton, N.J., to join the Center for Communications Research, the cryptography branch of the Institute for Defense Analyses, a nonprofit research center financed by the Department of Defense. His work earned him a name in the cryptography circle. "You can probably pick out two or three people who've really stood out, and he's one of them," said Alan Richter, a longtime scientist at the defense institute.

In 1993 Dr. Patterson moved to Renaissance Technologies, a $200 million hedge fund, at the invitation of its founder, James H. Simons, a mathematician and former cryptographer at the institute. The fund made trades based on a mathematical model. Dr. Patterson knew little about money, but the statistical methods matched those used in code breaking, Dr. Simons said: analyzing a series of data — in this case daily stock price changes — and predicting the next number. Their methods apparently worked. In Dr. Patterson's time with the hedge fund, its assets reached $4 billion.

By 2000, Dr. Patterson was restless. One day, he ran into Jill P. Mesirov, another former defense institute cryptographer, and mentioned his interest in biology. Dr. Mesirov, then director of computational biology at the Whitehead/M.I.T. Center for Genome Research, which later became the Broad Institute, hired him.

"Really, what we do for a living is to decrypt genomes," Dr. Mesirov said. Cryptographers look at messages encoded as binary strings of zeros and ones, then extract underlying signals they can interpret, Dr. Mesirov said. The job calls for pattern recognition and mathematical modeling to explain the data. The same applies for analyzing DNA sequences, she said.

One common genomic analysis tool — the Hidden Markov Model — was invented for pattern recognition by defense institute code breakers in the 1960s, and Dr. Patterson is an expert in that technique. It can be used to predict the next letter in a sequence of English text garbled over a communications line, or to predict DNA regions that code for genes, and those that do not.

Dr. Patterson said he also has a well-honed instinct about which data is important, after seeing "a lot of surprising stuff that turned out to be complete nonsense." Dr. Lander of the Broad Institute describes him as a great skeptic, with the statistical insight to tell whether a signal is "simply random fluctuation or whether it's a smoking gun."

Making that distinction is one of the great difficulties of interpreting DNA. In studying the human-chimp species split, the genomics researchers strove to rule out possible errors and biases in the data.

Dr. Reich, with Dr. Patterson and Dr. Lander, and two other colleagues, used computer algorithms to compare the primate genomes and count DNA bases that did not match, like the C base in gorillas that had become an A in humans. Because such mutations naturally arise at a set rate, the researchers could estimate how long ago the human and chimp lineages separated from an ancient common ancestor.

A DNA base can mutate more than once, however. To correct for that, Dr. Patterson worked out equations estimating how often it occurred; Dr. Reich revised their computer algorithms

accordingly. Two strange patterns emerged. Some human DNA regions trace back to a much older common ancestor of humans and chimps than other regions do, with the ages varying by up to four million years. But on the X chromosome, people and chimps share a far younger common ancestor than on other chromosomes.

After the researchers tested various evolutionary models, the data appeared best explained if the human and chimp lineages split but later began mating again, producing a hybrid that could be a forebear of humans. The final breakup came as late as 5.4 million years ago, the team calculated.

The project was "our hobby" Dr. Reich said of himself and Dr. Patterson said. Their main work, in medical genetics, includes devising a shortcut to scan the genome for prostate cancer genes.

Whether studying disease or evolution, Dr. Patterson noted, genomics differs from code breaking in one key respect: no adversary is deliberately masking DNA's meaning. Still, given its complexity, the code of life is the most open-ended of cryptographic challenges, Dr. Patterson said. "It's a very big message."