# The continuing evolution of DNA

## Some human genes take on essential tasks, and others fall to wayside

BY CARL ZIMMER

Each of us carries just over 20,000 genes that encode everything from the keratin in our hair to the muscle fibers in our toes. It's no great mystery where our own genes came from: Our parents bequeathed them to us. And our parents, in turn, got their genes from their parents.

But where along that genealogical line did each of those 20,000 protein-coding genes get its start?

That question has hung over the science of genetics since its dawn a century ago. "It's a basic question of life: How evolution generates novelty," said Diethard Tautz of the Max Planck Institute for Evolutionary Biology in Plön, Germany.

New studies are now bringing the answer into focus. Some of our genes are immensely old, perhaps dating all the way back to the earliest chapters of life on Earth. But a surprising number emerged more recently — many in just the past few million years. The youngest evolved after our own species broke off from our cousins, the apes.

Scientists are finding that new genes come into being at an unexpectedly fast clip. And once they evolve, they can quickly take on essential functions. Investigating how new genes become so important may help scientists understand the role they may play in diseases like cancer.

"It's premature to make any grandiose claims, but there's a coherence that's emerging," said David Begun, an evolution scientist at the University of California in Davis.

Geneticists first speculated about the origin of genes in the early 20th century. Some proposed that when cells duplicate their DNA, they accidentally copy some genes twice. At first the two genes are identical. But later, they evolve into different sequences.

At the end of the century, as scientists gained the ability to read the precise sequence of DNA, they found that this hunch was correct. "It became clear that gene duplication played a role in evolution," Dr. Tautz said.

As genes duplicate over millions of years, they can grow into so-called gene families, each containing hundreds of similar genes.

One family, for example, is essential for our sense of smell. These genes encode 390 different kinds of proteins produced in our noses, called olfactory receptors. Each olfactory receptor has a slightly different structure, allowing it to capture a different set of molecules.

Over long periods of evolutionary time, some copied genes change drastically — so drastically, in fact, that they take on entirely new tasks.

Consider hemoglobin, which stores oxygen in red blood cells for delivery throughout the body. Scientists have found that it belongs to a family of genes that do many different things with oxygen and recent studies suggest that it evolved from proteins that grabbed extra oxygen molecules inside cells before they could do harm.

The case for gene duplications became so strong that many scientists grew convinced that it was the source of all new genes. They speculated that when life originally emerged billions of years ago, the first primordial microbes had a tiny set of genes. Those genes then duplicated over and over again to give rise to all the genes on Earth today.

But when scientists gained the ability to sequence entire genomes, there was a surprise waiting for them. They started to find genes that existed in the genome of just one species. According to the duplication theory, these solitary genes shouldn't exist; they would have to have been copied from earlier genes in other organisms.

"They looked like perfectly normal genes, except they were only found in one species," said Anne-Ruxandra Carvunis, an evolutionary biologist at the University of California in San Diego. "There was no explanation for how a gene could be in one species and not in other ones."

These genes came to be known as "orphan genes." As scientists sequenced more genomes, they tried to return these orphans to their gene families. Sometimes they succeeded. But very often the orphans remained orphans.

For some scientists, like Dr. Tautz, the data pointed to an inescapable conclusion: Orphan genes had not been passed down through the generations for billions of years. They had come into existence much later.

"It's almost like Sherlock Holmes," said Dr. Tautz, citing the detective's famous dictum: "When you have eliminated the impossible, whatever remains, however improbable, must be the truth."

Dr. Begun and his colleagues renamed orphan genes "de novo genes," from the Latin for new. He found that many of his fellow scientists weren't ready to accept this idea. "It took a while for people to believe this was occurring," he said. "It seems kind of nutty to people when they first hear of it."

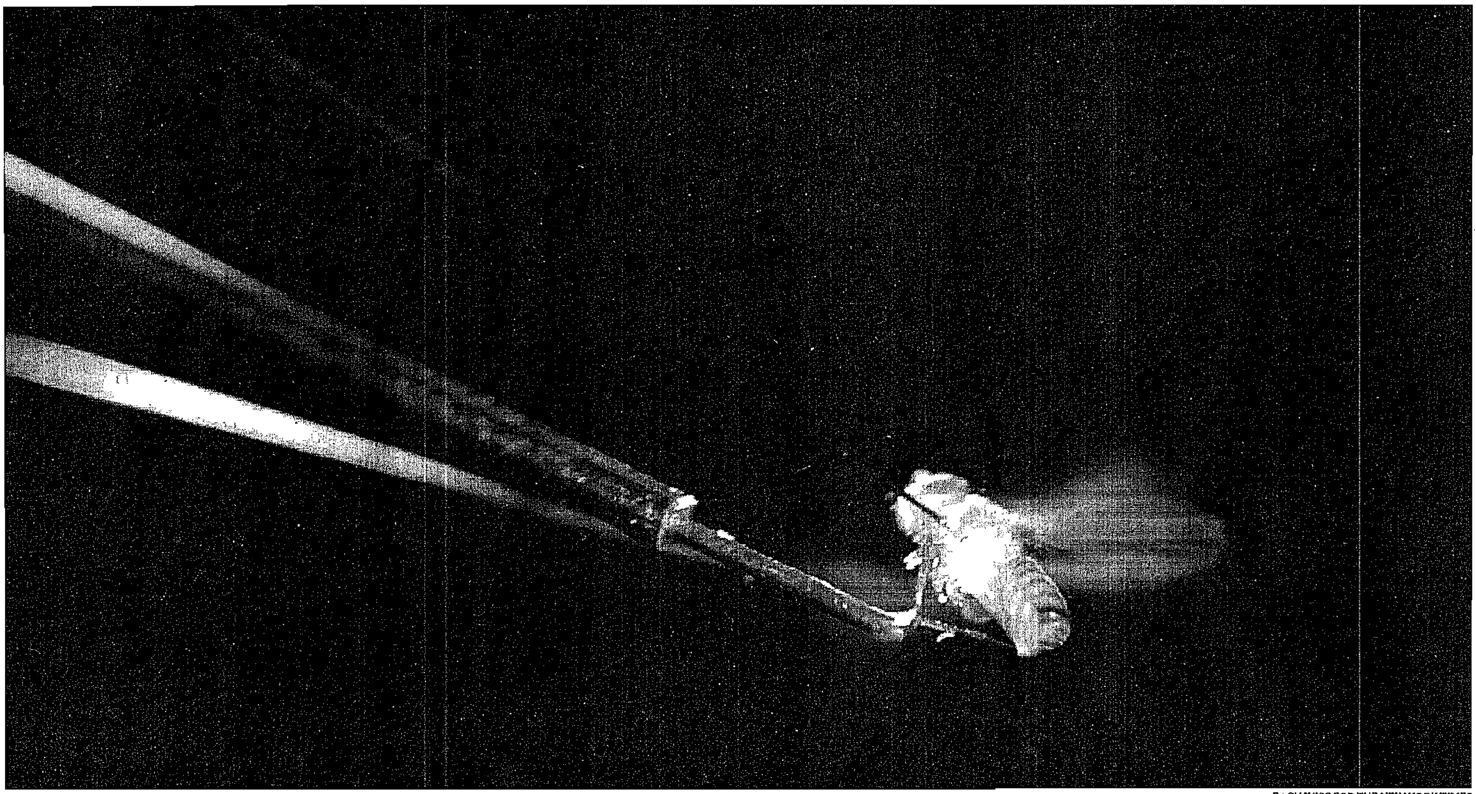
One reason it no longer seems so improbable is that Dr. Begun and other researchers have documented the step-by-step process by which a new gene can come into existence. In many species, ours included, protein-coding genes make up a tiny portion of the genome. New genes can emerge from the vast expanse of noncoding DNA. The first step is for a tiny bit of DNA to mutate into what scientists call a "start sequence." All protein-coding genes have start sequences, which enable cells to recognize where genes begin.

Once a cell recognizes the start of a gene, it can make a copy of the gene's DNA. It can then use that copy as a guide for building a protein. Once a new start sequence evolves, a cell will make a copy of the DNA next to it. The cell can then make a protein based on the new gene.

The new protein may turn out to be toxic, or it may just be harmless. But once it emerges, new mutations to the proto-gene may make it more useful.

Dr. Begun and his colleagues are now getting a look at these early stages in the birth of de novo genes. They look for such genes in different populations of a species of the fruit fly Drosophila melanogaster. They found 142 de novo genes that were present in some populations of flies and not in others, meaning that they must have evolved recently: They've only had enough time to spread across part of the species.

Dr. Begun suspects that the true number of de novo genes in the flies is higher. He and his colleagues used very strict guidelines about what stretches of DNA they put on their list, and so they may have missed some genes.



ZACH WISE FOR THE NEW YORK TIMES

**The fruit fly Drosophila melanogaster is a model for the study of human genetics. Some research has indicated that new genes can evolve in the species at a remarkably fast rate.**