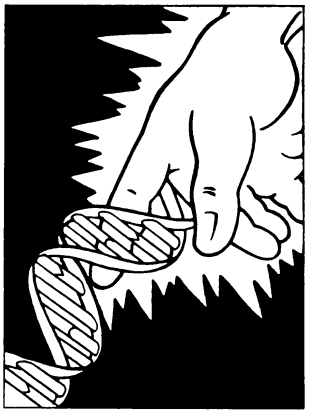


In Search of Genesis and the Pseudogene

A pre-med student at Walla Walla College begins a quest he still pursues as a fourth-generation Adventist on the Harvard Medical School faculty.

by Gary Gilbert

FROM THE BACK OF THE BUS I GROANED AS another lecture began. We had been traveling all night and, at 6 a.m., were whistling along a highway in southern Utah. "If you will look at the sedimentary rock formations on our right . . ." we were directed. The heavy silhouette of gray rock had taken a pinkish hue in the dawn light and the lines, the boundaries between sedimentary layers, were barely distinguishable. The lecturer, a geologist, began to tell us about the fossils in the formation we were passing. I rubbed my eyes as the speaker explained that fossils within the mass of rocks were a record of past life at a time when this part of Utah was underwater. A college freshman and a fourth-generation Seventh-day Adventist, I started the trip believing that apparent problems with the Genesis story could be solved if you were armed with a knowledge of Noah's flood and an open mind. Influenced by my father, who was a physicist, I did not believe that scientific findings could be ignored or trivialized; rather, I believed that God was responsible for both the natural world and for Scripture. Nature spoke both about itself and about God. Scripture gave advantageous insights to Christians, a head start in the study of nature, but its authors were not scientists and neither they nor their modern interpreters should have the last word when nature spoke clearly. The effort of our field trip guides was to understand the fossils and the many rock layers in the context of a short history of life. If not 6,000 years, perhaps 12,000. As we gazed at the wall of the Grand Canyon, turned fossils over in our fingers, and discussed the explanations in an open air Sabbath school, I came to realize that sedimentary layers of the Southwest could not all be explained by a recent Creation and a great Flood. Around the campfire I heard whispers that the two SDA geologist guides did not agree with each other about the type of natural events implied by sedimentary layers that we had seen. I can't remember the

Gary Gilbert, who teaches at Harvard Medical School, directs a hematology research laboratory. He received his M.D. from Loma Linda University Medical School. *Gilbert wishes to express gratitude to Sandra Robb-Gilbert for persistent revisions that make this piece more accessible to the non-scientist.*

substance of the dispute or the clues that made my classmates aware. I can remember the dawn of my awareness that the story the rocks told was of many floods, not just Noah's flood, occurring over a time span of much more than 6,000 years.

Two years later, in an upper-division religion course designed for science majors, we discussed areas in which religion and science were in conflict. We discussed various methods for determining historical age: dendrochronology, carbon-14 dating, potassium-argon dating, and dating based upon scriptural genealogies. I remember the simplicity and clarity of dendrochronology and was impressed that if one counted the consecutive rings from living fossilized trees that had grown in that same grove, time marched backward right past the date for Noah's flood. Was the date wrong or did the Flood not uproot the trees? Again, the age of life on earth, as indicated by geological and physical methods, was at the center of the discussion.

What did not seem strange, at the time, was the paucity of serious discussion about evidence that animals of one type have evolved from animals of a quite different type. The most thoughtful Adventist scientists that I knew were concerned with the age of life more than the ancestry of living animals. Darwin's hypothesis, though viewed as a threat to Adventist beliefs, was not the center of the seminars or discussions. Darwin's hypothesis, that animals are related to one another by common ancestry, is not primarily about time. It is about the mechanism through which animals acquire new characteristics and change dramatically over generations.

Molecular genetics was new, slow, and cumbersome at the time of my entrance to La Sierra University in 1972. More recently, molecular genetics has moved into the fast lane, becoming the primary basis for experimental biology and for the multi-billion dollar biotechnology industry. Twenty years ago, molecular genetics offered a fresh opportunity for creationists to find support for a brief duration of life upon earth and for the separate genetic lineages of different "types" of animals. Instead, the emerging genetic information supports Darwin's hypothesis—that animals are related to one another by descent from a common Creator. Unlike older fossil evidence, the genetic data is not dependent upon the estimated age of the earth. If the Grand Canyon and all of the sedimentary layers in the Southwest can be explained by a series of catastrophes occurring over only 6,000 or 12,000 years, the genetic data will still suggest that most animals are descendants of a common ancestor. I will recount my exploration of genetic findings that illuminate the most vehemently contested relationship, the one between great apes and humans.

Mistakes Are Best Explained by History—Not Teleology

General biology was a prerequisite for medical school, and half of my 700 college freshman classmates planned to enter medical school. As a result, all three sections of Biology 101 were crowded. Every time the professor spoke in his faded British accent, 90 pens scratched on note paper. One morning he lectured about protein molecules of humans and animals that were almost identical. With apparent disdain, he noted that some people argued that molecular similarity supported evolution. If God worked out a good design for hemoglobin once, he asked, why wouldn't he use the same design again when he created humans? Caught up in his debate against an absent adversary, he demanded to know what happened to machines left alone. Did they become more complicated, more excellent? No. Then how could the evolutionists propose that neglect and chance made animals become better and more complicated,

generation by generation? Ninety pens scratched on note paper. No one was taking chances about what might be on the quiz. Six months later, during the spring quarter, the aura of intense determination had faded from Biology 103 and, as the grades accumulated, many classmates admitted that they might never enter medical school. I was daydreaming about a summer visit to see my girlfriend during biology lectures in a classroom that now had many empty seats. My professor's arguments about similar molecules not suggesting evolution had lodged firmly in my mind—ready to prevent further questioning about molecules and evolution for 15 years.

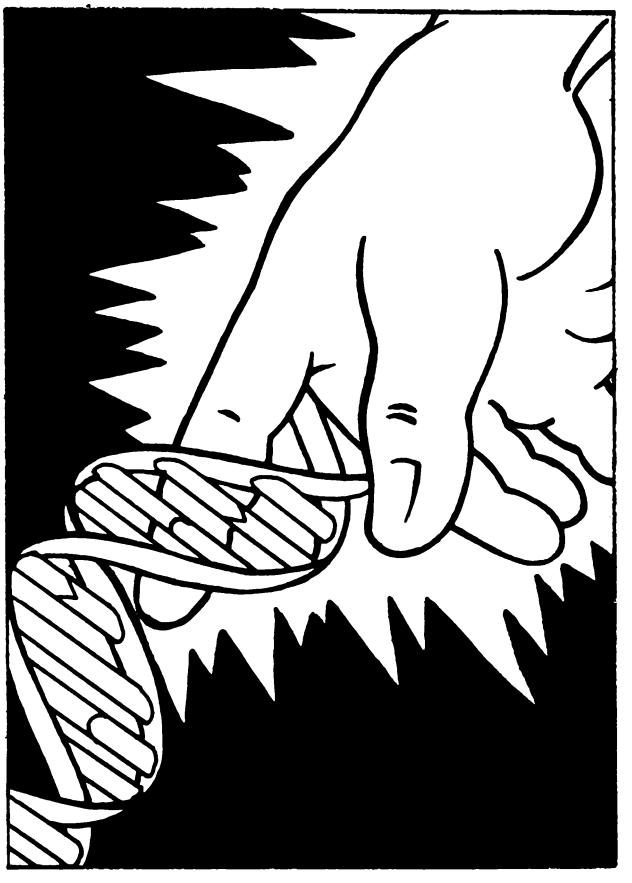
When I entered a research fellowship at Tufts University–New England Medical Center after medical school and residency, molecular genetics had transformed the way that biology was studied. In seminar after seminar a new human gene would be described and compared to a similar gene in a mouse or a cow or a yeast. I became aware of the pervasive genetic similarity between animals. If a scientist wanted to identify a new human gene and the human tissue in which the gene functioned was difficult to obtain, a reliable way to identify the human gene was to find it first in another mammal. The genes of different animals were not absolutely identical to one another. Genes were depicted as a long string of letters (only the letters A, C, G, and T were used). If the letters from a human gene were aligned with the letters from a cow gene, about 70 percent of the letters would be identical. I gradually became aware that this pattern was not what my professor's explanation for the similarity between human and animal proteins predicted. If God used the same plan for hemoglobin protein when he made cows and humans, then the hemoglobin proteins should be identical—or any differences between them should serve a purpose. The cow hemoglobin carried oxygen in just the same way that human hemoglobin did. In some cases, the evidence that the differences between the human and the animal protein did not serve a purpose was simple—the animal protein functioned normally in a human.[1] In other cases, the purposelessness of the differences could be surmised because biochemical studies indicated that the animal and human proteins functioned equivalently in spite of a few differences. No single example changed my understanding of creation. Gradually though, I became aware that if God, like a good engineer, had used a single genetic design for a protein in different animals, then the quality control on his production line was poor.

No single example changed my understanding of creation. Gradually, though, I became aware that if God, like a good engineer, had used a single genetic design for a protein in different animals, then the quality control on his production line was poor.

It was not the human genes that were similar to animal genes that finally focused my attention on genetics and Creation. At an early-morning science seminar—scientists consider 8:15 early for a seminar although doctors do not—Dr. Sadler told us about the pseudogene that he had discovered, quite by accident.[2] The particular pseudogene was nearly identical to the gene that encoded the protein named von Willebrand factor. A pseudogene, I learned, is a flawed copy of a

gene. The flaw is sufficiently destructive that the pseudogene cannot possibly function as a gene. Apparently the result of a rare type of genetic mistake, a pseudogene may become an integral part of the genetic code. Like a gene, it is carried by all offspring of the individual in whom the genetic error occurred. In contrast to a gene, the pseudogene has no effect upon the person or animal who carries it. While it may be surprising that an extra gene could have no effect, it is apparently true. Because only 2 percent of the DNA in humans has a specific function—98 percent is apparently silent—an extra pseudogene may "go along for the ride" without causing impairment. The von Willebrand factor pseudogene had so few DNA letter differences from the authentic gene that our speaker predicted that it might have occurred recently in the course of evolution; perhaps after the human and chimpanzee species divergence from monkeys. I was entranced. Not by the possibility that the pseudogene might not be carried by monkeys but the possibility that it was carried by chimpanzees. Before me was a genetic marker whose presence in different animals would unambiguously indicate common heredity for those animals. Because this genetic marker had no function, there was no motive for a good designer to include it in the design of different types of animals as they were created. Therefore, its presence in different animals could only be explained by a common ancestry, not by the actions of God as designer or engineer.

Dr. Sadler is known for efficient work. I estimated that it would take him 18 months to search for the von Willebrand factor pseudogene in chimpanzees and gorillas and report his results. If the pseudogene were in chimpanzees it would be strong evidence for common ancestry of humans and chimps. I didn't want to wait 18 months for an answer and realized that there must be other pseudogenes. Perhaps a pseudogene shared by humans and chimps had already been identified. I began to spend additional time in the library reading about pseudogenes. The information I was looking for was scattered because molecular geneticists have long believed that humans and chimpanzees share ancestry and neither the titles of the papers nor the discussions emphasized this point. I learned quickly that, genetically speaking, humans and chimpanzees are almost identical in all genes that have been decoded.[3] If a string of letters for a human gene is placed side by side with the string of letters for a chimpanzee gene, differences are found less frequently, on average, than one out of every hundred letters. Genetic differences occur more frequently between animals many people would see as more closely related. For example, differences between the genes of various sea urchin species from different ocean regions occur much more frequently than between chimps and humans. In another comparison, the genes of mice and



rabbits differ from each other 20 times more frequently than do those of humans and chimps.

During my pursuit of pseudogene reports, the idea of using errors to identify sources permeated my thinking. A literary example appeared in a *National Geographic* article about Christopher Columbus: a book once owned by the explorer was filled with marginal notes. How did a scholar 500 years later determine whether those notes were really scribbled by Columbus? A clue came from Columbus' imperfect use of Latin. The notes contained Latin errors similar to those made by Columbus in other authenticated manuscripts. It was not the appropriate note, even the characteristic thought, that was so compelling in identifying Columbus as it was the source of his characteristic errors.

I thought of another example. Suppose that I was examining manuscripts that were stored on the hard disk of my computer, and I discovered that the second and third paragraphs of a letter to my city tax abatement board were inexplicably trailing at the close of a book review I was revising. The wording in those paragraphs was strong and clever (at least I thought so) and was identical with the original letter to the tax abatement board. Furthermore, when I examined another copy of the book review on a floppy disk, I discovered that the same paragraphs were attached to the back-up copy of my book review. I would conclude that the paragraphs from the letter requesting tax abatement were copied by mistake—either mine or the computer's—into the book file; but that once the mistake was made on the hard disk, the computer faithfully copied the whole manuscript, with the mistake, onto the floppy disk. I would not conclude that the recording space on the floppy disk spontaneously changed in a way that gave it such a close resemblance to my original letter or the book review. I would be able to identify copies of the book review that were made before or after the error by finding whether they lacked or included the mistakenly appended paragraphs—to identify copies made from the "genetic code" that included it.

It did not take many days in the library to identify reports of pseudogenes that were present in both the human and chimpanzee genetic code. The first examples that I found did not satisfy me. I was not convinced, in several cases, that evidence showing a particular pseudogene had no function was conclusive. Then one morning in another seminar the speaker made reference to a hemoglobin pseudogene reported six years earlier. The same afternoon I located the papers reporting the sequence of this pseudogene in humans and chimpanzees. It was the compelling example that I had been looking for.

The β Hemoglobin Pseudogene

There it was, a genetic signature left by an ancestor of mine . . . an ancestor that I share with chimpanzees. I sat quietly for at least an hour comparing the DNA letters of the human pseudogene and the chimpanzee pseudogene. This pseudogene, called the β hemoglobin pseudogene, is large and located next to the corresponding normal β hemoglobin gene.[4,5] The chimpanzee β hemoglobin pseudogene is also located next to the same normal hemoglobin gene.[6] The table on page 15 shows some of the DNA letters that I compared. All of the first 63 letter are identical in the human and chimp pseudogene, indicated by the symbol " below the corresponding letter of the human gene. Only six DNA letters of the chimp pseudogene differ from the corresponding letters of the human pseudogene out of a total of more than 500.[7] The six DNA letters that are different in the human and chimpanzee pseudogenes are

believed to be the result of random mutations that have occurred in both pseudogenes since they were inherited from a common ancestor.[8] Random changes in single DNA letters occur at a slow but predictable rate over generations. The number of single letter differences between the human and chimpanzee pseudogene suggests, using the "molecular clock" technique, that the common ancestor lived between 4 million and 6 million years ago.[9]

I was satisfied that this pseudogene was really a functionless segment of DNA. It was clear that the β hemoglobin pseudogene could not function as the plan for protein. The bottom row in the table below indicates the "meaning" of the DNA letters in the pseudogene. There is only one sequence of three DNA letters, ATG, that can mark the beginning of a protein to be synthesized. The beginning signal in the A hemoglobin gene, from which the β hemoglobin pseudogene originated, is changed in the β hemoglobin pseudogene. The symbol in the bottom row ~~START~~, indicates this problem: no start signal. Even if the β hemoglobin pseudogene had a start signal, the hemoglobin formed would stop prematurely at the position of the 15th amino acid, which has the letters that indicate STOP. There are several additional STOP signals throughout the β hemoglobin pseudogene, further eliminating the possibility that the pseudogene could function as a gene.

Further evidence that the β hemoglobin pseudogene is really a pseudogene is our ability to identify the functional gene from which it was copied and whose function it now lacks. The resemblance to the A hemoglobin pseudogene is illustrated in the table. The vertical lines between the letters in the upper and lower rows, corresponding to the A hemoglobin gene and the β hemoglobin pseudogene, are a visual aid to identifying the DNA letters that are identical. In the displayed region, 44 of 63 DNA letters in the sequence, or 70 percent, are identical. For the entire gene (about 500 DNA bases) the fraction that are identical is also about 70 percent. For comparison, imagine two unrelated segments of DNA aligned in this way. You would find that approximately 25 percent of letters were identical. The probability that a random process would lead to 70 percent similarity over a DNA

Comparison of the Human Aγ Globin Gene With the Human and Chimp β Hemoglobin Pseudogenes

```
Human Aγ Gene  G C C A T G G G T C A T T T C A C A G A G G A G G A C A
                 |     |   |       | | | | | | |   |   | | |     |   |
Human β        A C T G T A G T G C A T T T C A C T G C T G A C A A G A

Chimp β        " " " " " " " " " " " " " " " " " " " " " " " " " " " "

Amino Acid Sequence  START   Val   His   Phe   Thr   Ala   Asp   Lys
(β hemoglobin pseudogene)

A G G C T A C T A T C A C A A G C C T G T G G G G C A A G G T G A A T
| | | | |   | |     | | |   | | | | | | | |     | | | | | | |   | | |
A G G C T G C T G C C A C C A G C C T G T G A A G C A A G G T T A A G

" " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " " "
Lys Ala   Ala   Ala   Thr   Ser   Leu   STOP Ser   Lys   Val   Lys
```

(In the printed table, "START" in the amino acid row is struck through.)

chain 500 units long is nil. The 30 percent of DNA letters that are different are believed to be the result of random mutations that have occurred in both genes since the original flawed copying of the A hemoglobin gene.

The most compelling evidence that the pseudogene has no function is that people do not need it for good health. Genetic errors causing faulty hemoglobin synthesis have been identified in hundreds of patients, yet none is traceable to defects in the β hemoglobin pseudogene. There are people who lack part of the β hemoglobin pseudogene because of a genetic mutation. These patients do have abnormal hemoglobin but it is entirely explained by loss of the adjacent hemoglobin gene. No problem can be attributed to living with an incomplete β hemoglobin pseudogene. Together, this information convinced me that the β hemoglobin pseudogene arose from another hemoglobin gene, that it does not function as a hemoglobin gene, and that it lacks any function whose absence would cause a health defect.

Subsequently, I found reports of other pseudogenes. There is a pseudogene in the (a) cluster of hemoglobin genes that is also shared by chimps and gorillas. There are several probable pseudogenes in the gene complex that codes for immunity recognition molecules, and there are others. Some of the probable pseudogenes in the complex codes for immunity recognition molecules are also present in the chimp genetic code. The existence of all of these pseudogenes supports the same idea—that humans and chimps share a common ancestor.

Alternative Explanations

I felt anxious after my discovery. I worried that I had jumped to a conclusion while overlooking the alternatives. It was possible to imagine other explanations for shared pseudogenes than common ancestry. I reexamined the explanations that I knew and discussed them with other Adventist scientists. For example, suppose that the β hemoglobin pseudogene really does have a function and is not just a flawed copy of a hemoglobin gene. If so, then presence of the pseudogene in other primates could conceivably be explained by use of a common gene design by the Master Designer, rather than by common ancestry. If there is such a function, it is not as a gene. The β hemoglobin pseudogene has STOP signals too frequently for this. Recent technological advances with transgenic animals and with embryonic gene insertion make it feasible to design an experiment to test the hypothesis that the β hemoglobin pseudogene has an important function. Perhaps an Adventist graduate student, convinced that the β hemoglobin pseudogene has an important function, will risk three years of her graduate program to perform the appropriate experiments . . .

It may be tempting, in view of the 19th-century evangelical beliefs about "amalgamation of man and beast," to speculate that the shared pseudogenes may be a result of interspecies breeding between humans and apes.[10] This line of reasoning would require that all humans containing genetic markers common to apes be descendants of human-ape breeding (this includes all of the thousands of humans studied to date); not only those with dark skin as some 19th-century writers believed. Statistical comparison of gene differences between various primates suggests that common ancestors may have interbred sporadically, but not within the past 5 million years.[11]

Genes may be transferred between individuals by viruses in the laboratory. This is referred to as "lateral gene transfer" and is the basis for gene therapy that is now being tried in humans. I did not believe that this was a likely explanation for the pseudogenes, however. First, under natural conditions the proc-

ess is largely unknown. A baby gets type A blood because the genes for type A were inherited from her parents, not because her mother contracted a virus infection from a friend with type A blood during conception or pregnancy. The viruses that have the capacity to carry genes as hitchhikers do so under contrived laboratory conditions. The possibility that cancer-causing human genes are transmitted by viruses was once a favored hypothesis, but after several decades evidence that this is a mechanism for transmitting human cancer causing genes from person to person is still lacking. Second, the location of genetic material inserted by a virus is random, while the β hemoglobin pseudogene is always found adjacent to the normal β hemoglobin gene in humans, chimps, and gorillas. Third, viruses insert their own viral genes adjacent to any mammalian gene that has been carried along. Thus, viral genes in the human genetic code serve as markers of the nature and location of the gene acquisition. Viral genes have not been found adjacent to the β hemoglobin pseudogene.

It remains theoretically possible that the mutations that have led to pseudogenes have occurred independently in different animals. The small probability of this may be grasped by thinking back to the example of the inexplicably copied paragraph from the letter to the tax abatement board. If the faulty appended paragraphs were identical in two copies of the manuscript recorded on two different floppy disks, I would conclude that the mistake had occurred once, and then the manuscript duplicated in the normal way. It is far less likely that the rare event of flawed gene copy and insertion occurred at the same time, with the same amount of copied material, and in the same place, in chimpanzees, humans, and gorillas.

In the end the alternative explanations all seemed contrived to me. I also recalled that pseudogenes had not been the first genetic evidence that had suggested common ancestry rather than common design. The other evidence was substantial on its own merit. Although pseudogenes may have been capable of standing alone as an elegant proof for common ancestry of humans and chimps, for me their evidence was confirmatory. The did not stand alone.

Reflections

The thrill of a new insight was tempered by a sense of loss as I contemplated the β hemoglobin pseudogene. While I had long assumed that parts of the first chapter of Genesis spoke metaphorically (there was no other option after studying physics), the description of God forming a clay model for Adam followed by suffusion with life had not stimulated my doubt. My next reading of Genesis left me sad, for I felt a little closer to the animals and a bit farther from the Sculptor who wished to make us in his image.

My knowledge of geology has increased sporadically, and in small increments, since my college field trip. Reports in journals such as *Science and Nature* discuss an age for life on earth estimated in hundreds of millions of years, not a few thousand years. The Adventist claim that contradictory data prevents scientific consensus about the age of life on the earth is not supported by reports in these widely read journals.[12] I wonder if freshman students at La Sierra University still take geology field trips to Utah led by guides who are struggling against a barrage of scientific reports to reinterpret fossil findings in terms of Noah's flood and other events occurring over a few thousand years.

In college I found it easiest to dismiss the scientific techniques and evidence that I understood the least well. I still do. Adventist colleagues with whom I have discussed the ancestral link between people and chimps

implied by genetics have a background, like my own, in which center stage in the conflict between Adventist creationism and science was previously held by the age-of-life question. The age-of-life question is no longer an issue with which they struggle, having been resolved in favor of epochs much longer than 6,000 years. The issue of human ancestry is receiving increasing attention. While the mechanisms of molecular genetics are familiar to anyone trained in biological sciences within the past two decades, the data supporting common ancestry for humans and great apes is not widely known among Adventists. Because common ancestry has long been considered established by the scientific community, the genetic findings that confirm common ancestry are not emphasized in scientific journals. Those Adventists who are familiar with this information seem unable to dismiss it, and exploration of the implications is apparently ongoing.

I am curious about the eventual impact of molecular genetic findings upon Adventist creationism. Because of the broad utility of molecular genetics to a burgeoning biotechnology industry, more Adventists will learn this discipline than geology. Perhaps impetus from these Adventists will lead to a re-examination of acceptable interpretations of the ancient Hebrew document, Genesis. I suspect it is more likely that those who understand genetics—and care about a synthesis between the world they study and their religious faith—will continue to limit discussion to discreet conversations amongst themselves. The outcome may be influenced by what Adventist college freshman now hear during lectures on biology and geology. It would be interesting to audit a biology class at my alma mater to see if the implications of molecular genetics have filtered into the curriculum. I would listen for a hint of a fresh Adventist approach that neither distorts the science of genetics nor equates reasoned interpretation of genetic data with abandonment of faith.

NOTES AND REFERENCES

1. Insulin is an example of an animal protein in which slightly different genetic sequence and resulting protein structure have not prevented an animal protein from functioning normally in humans. Until 1983, all diabetics received insulin that was extracted from slaughterhouse products of pigs or cows. In 1983, human insulin, produced by genetically engineered bacteria containing the human insulin gene, was first marketed for human use by Eli Lilly and Company.

2. A full report of this work was subsequently published under the title *Human Willebrand Factor Gene and Pseudogene: Structural Analysis and Differentiation by Polymerase Chain Reaction* by D. J. Mancuso, E. A. Tuley, L. A. Westfield, T. L. Lester-Mancuso, M. M. Le Beau, J. M. Sorace & J. E. Sadler, *Biochemistry* 30 (1991): 253-269.

3. While less than one percent of the common genetic codes of humans and chimps have been determined, many thousands of DNA letters are available for comparison.

4. For a review of the molecular genetics of hemoglobin, both the contribution to understanding of human disease and the interpretation of genetic information to derive hereditary lineages, see the textbook by H. F. Bunn, *Hemoglobin—Molecular, Genetic, and Clinical Aspects* (Philadelphia: W. B. Saunders, 1986).

5. Because there are diseases caused by defects in hemoglobin protein, such as sickle cell anemia and thalassemia, these genes have been carefully and frequently analyzed to identify the causative defects. A form of gene analysis is now part of prenatal testing for couples concerned that their child may have sickle cell anemia.

6. The entire sequence of the β hemoglobin pseudogene for human, chimpanzee, and gorilla are contained in the article by L. Y. E. Chang and J. L Slightom, "Isolation and Nucleotide Sequence Analysis of the β-type Globin Pseudogene From Human, Gorilla and Chimpanzee," *Journal of Molecular Biology*, Vol. 180, pp. 767-784.

7. The number of letters (bases) to include in the β hemoglobin pseudogene was chosen, somewhat arbitrarily, as the number with homology to the coding portion of the Aγ gene. The sequences corresponding

to Aγ introns have not been counted.

8. The similarity between the functional hemoglobin genes of humans and chimpanzees is even closer than the similarity between pseudogenes. That is, there are even fewer differences in the sequence of DNA bases that code for the globin proteins.

9. Although a time scale is not required to support the evidence for common ancestry between humans and chimps, the time since descent from a common ancestor can be estimated if a constant genetic mutation rate is assumed—see under "A Primer on Molecular Genetics," below.

10. This possibility has been suggest by Dr. L. J. Gibson, a member of the Geoscience Research Institute. He suggests that genetic "sequence may have been transferred from one species to another by *introgression* . . ." (italics supplied). The term *introgression* indicates a hypothesis that pseudogenes were transferred via cross breeding between species. That is, humans, chimpanzees, and gorillas (species that share the β hemoglobin pseudogene) have interbred freely to the extent that the shared genes are present in all members of each species. This is a peculiar hypothesis to be promoted by the Geoscience Research Institute, since the proposed interspecies breeding would provide a simple mechanism for interspecies evolution, a process they argue has been rare or nonexistent. See L. J. Gibson, *Dialogue*, 3:36 (1991).

11. M. Hasegawa, H. Kishino, and T. Yano, "Man's Place in Hominoidea as Inferred From Molecular Clocks of DNA," *Journal of Molecular Evolution*, Vol. 26, pp. 132-147.

12. A recent example of this claim is in *Dialogue* (Vol. 2, No. 2), the Adventist journal produced for Adventist academics. Dr. L. J. Gibson asserts that scientists who believe that life came into being 600 million years ago have about the same amount of data supporting their beliefs and about the same number of obstacles to surmount in supporting their theory as those who believe life originated 6,000 years ago. He implies that one who reads scientific reports is likely to encounter as many competent reports from researchers that believe that 6,000 years is a good estimate to those who estimate 600 million years.

Appendix A
A Primer on Molecular Genetics

DNA stands for deoxyribonucleic acid, a long molecular chain composed of four types of "chain links" that carry genetic information from one generation to the next. From the double helix structure scientists learned that genetic information is carried in the discrete "chain links" of the DNA. The four types of "chain links" are bases,[1] referred to by the letters A, T, C, and G. Information is carried by the sequential order in which they are arranged in the chain. In the same way that two symbols, 0 and 1, carry information about language, shapes, colors, and actions to a computer, based upon their arrangement in a long string of symbols, the arrangement of the four DNA bases carries information specifying a human. This means that every inherited trait a person possesses—characteristics such as skin color, height, athletic ability, et cetera—may be ultimately traced to a series of DNA chains, base by base, has become widely available in the last decade so that now any graduate student can determine the arrangement of thousands of DNA molecules that may determine a specific human trait. A large national project is underway to "read" all of the DNA information present in a human being, about 3 billion units (the human genome) together with variations that determine differences between individuals. This genetic information, which is simultaneously becoming available for plants and animals, is changing traditional disciplines. Molecular genetics now has the last word in phylogenetic classification of animals, is in frequent use for identification of criminals from small bits of tissue, and has been used for resolution of parenthood in disputed cases.

The information in DNA chains functions to specify construction of proteins. It is these proteins that carry on the business of life. For example, hemoglobin—a transporter protein—carries oxygen from the lungs to body tissues; trypsin—an enzyme—cuts food proteins into pieces so that they may be absorbed from the intestines; and myosin—a contractile element—ratchets along another protein after receiving a nerve signal, providing the muscle force to walk or run. All of these proteins are chains of amino acids, and every amino acid in the chain is specified by a corresponding DNA base sequence.

It is useful to think of the genetic code as a very simple language (see box, page 15).[2] In this lan-

guage the alphabet has only four letters—A, C, T, and G (each letter corresponds to one of the four types of DNA bases). There are no punctuation symbols such as spaces, commas, or semicolons; only START at the beginning of a long "sentence" and STOP at the end. Information is carried as a long string of letters such as ATTCGTCCA, et cetera. Like the English alphabet, the genetic alphabet is used to spell words, but the spelling rules are much simpler. Words contain three letters so that the string of letters above could be thought of as three words with spelling ATT, CGT, CCA. With only three letters per word and only four letters in the alphabet, it is clear that there are not very many words in the genetic language. If you like numbers you have probably already figured out that the genetic language has only 64 spellings for words. Each letter in a word can have four values, and there are three letters, so that the possible combinations are $4 \times 4 \times 4 = 64$. But the genetic language does not need 64 words; it requires only 21. The meaning of each "word," as it is translated into a growing protein, is an amino acid. There are only 20 amino acids (the building blocks of protein), plus the meaning STOP, specifying the end of an amino acid chain.[3] This leaves 43 extra spellings after 21 are claimed for the key functions. Extra spellings provide alternates for the 21 meanings so that 18 of 20 amino acids and STOP can all be spelled more than one way. For example, STOP is spelled TAA, TAG, or TGA; and tyrosine, an amino acid, is spelled TAT or TAC.[4]

A gene is a segment of a DNA chain that contains all of the information to make one protein. If the genetic code is modeled as a language, then a gene is a "sentence." It is a long chain of DNA bases that have the code for START (the DNA base sequence is ATG) at the beginning and the code for STOP (the sequence is TAA) at the end. In addition to genes, DNA chains contain other segments with other functions. For example, sequences of DNA located close to genes respond to a molecule that carries the message "get ready to make a protein." These sequences function much like an on-off switch. Other sequences are involved in DNA replication. Most surprising is the finding that large segments of DNA do nothing at all. Current evidence indicates that as much as 98 percent of the human genome may have no regular function. Large segments of DNA lack the START and STOP signals necessary for making proteins; therefore they are not genes. They lack the patters that are used for functioning as on-off switches or for involvement in DNA duplication. While geneticists believe that some of the silent DNA has had a critical role in evolution over hundreds or thousands of generations—related to rare, useful genetic mutations—it apparently has no specific impact over the life span of an individual.

Genetic Mistakes Allow Lineage Determination and Time Estimates

Every time a human cell divides it faces the formidable task of copying all DNA chains with 3 billion bases of information. It must provide a copy for each of two daughter cells. Skin cells, blood cells, brain cells, and fertilized ova all carry the same genetic information and must duplicate it during growth. What happens if a cell makes a mistake in duplication of the genetic code? The simplest mistake involves a single DNA base substitution for another base. There are three possible results of this type of mistake. It is most likely that the new "word," resulting from the change, will specify a different amino acid, and a modified protein will be synthesized. For example, if the second letter in a word specifying glutamic acid, GAG, were changed to C, then the new word would be GCG, which specifies valine, a different amino acid (see box, page 15). When this mutation occurs in the sixth word specifying the A chain of hemoglobin it causes sickle cell anemia.[5] The second possibility is that the new "word" may specify STOP. If the first letter of CGA, specifying arginine, is changed to T, the resulting word, TGA, causes synthesis of the protein to terminate at this word rather than adding an arginine to the growing protein and continuing the synthesis. This mutation has occurred at "word" number 2,307 of the factor VIII gene—causing hemophilia. It is also possible that the new "word" will be an alternate spelling for the amino acid originally specified. If the first T in TAT, specifying tyrosine, is changed to an A, resulting in ATC, the new word specifies tyrosine and now the mutant DNA will specify exactly the same protein! All of us carry genes specifying normal proteins that have alternate spellings.

The vast majority were inherited from our parents rather than occurring *de novo* in our own cells. This type of mutation is useful for tracing heredity in disputed paternity cases and in identification of criminals from tiny bits of tissue.

Again, if you like numbers, you may have already guessed that a modified protein is the most likely outcome for a randomly changed letter in the genetic language. The same protein specified by a modified code is about 15 times less

likely. Substitution mistakes occur at the rate of about 3 per cell division (or 1 per billion DNA bases copied).

The rate at which random mistakes in DNA accumulate is similar in many types of plants and animals. Because the mutation rate is constant, the elapsed time since an ancestor was shared by two populations can be estimated by counting the number of randomly distributed differences in a DNA segment that is otherwise identical. This method for estimating elapsed time is termed the "Molecular Clock." It is best applied to segments of DNA that have no function. This eliminates the bias introduced by mutations that cause a disease when present in a functional gene (such as sickle cell anemia). Defects in a functional gene may cause a survival disadvantage to the recipient and result in the accumulation of fewer mutations than anticipated based upon the random mutation rate.

A rare genetic mistake involves making an extra copy of an entire segment of a DNA chain and inserting it into another place in the DNA chain where it does not belong. This type of mistake occurs so infrequently that it is difficult to study in the laboratory. If a "sentence" or gene is copied, the new (and extra) gene probably will not function. It may be copied incompletely, lacking the START signal, or be truncated before the STOP signal. It may lack the nearby control sequences necessary to turn it on as a gene. The nonfunctioning "sentence" is called a pseudogene with reference to the gene from which it was imperfectly copied. Once the mistake has occurred, however, it will be transmitted to all cells that are offspring of the mutant cell. The vast majority of mistakes affect only a small number of cells—for most cells divide only a few times before dying. In order for a mistake to be passed to a baby in the next generation, it must occur in a germ cell—one that will become an egg or a sperm.

NOTES AND REFERENCES

1. While the A in DNA stands for acid, under biological conditions hydrogen ions are dissociated from the individual acid units, hence they are referred to as bases.

2. For a more thorough but still readable introduction to molecular genetics, see B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, J. D. Watson, *Molecular Biology of the Cell* (New York: Garland Publishing, Inc., 1983), pp. 98-110.

3. The meaning START does not have a separate spelling but is identical to the spelling specifying the amino acid methionine.

4. You may have noted that any string of DNA bases can be read three different ways, depending on which based you start with. The sequence ATTCGTCCA, for example, may be read as ATT, CGT, CCA or . . . A, TTC, GTC, CA . . ., or . . . AT, TCG, TCC, A . . . The way a particular sequence is recognized by the cellular mechanism to guide protein synthesis is referred to as the "reading frame." Because the signal STOP will occur about once in every 20 DNA words in a randomly arranged DNA sequence, one method for identifying a gene, and the correct reading frame, is to search for a long segment of DNA without a STOP signal. It is possible, in theory, to have overlapping genes in different reading frames. This is common in viruses where efficiency is at a premium, but it is very rare in mammals.

5. In most cases of sickle cell anemia this mutation is inherited from parents, but the original case arose, and occasional new cases stem, from new mutations.