

# GENERAL SCIENCE NOTES

## PSEUDOGENES AND ORIGINS\*

By L. J. Gibson, Geoscience Research Institute

### WHAT THIS ARTICLE IS ABOUT

*Pseudogenes are DNA sequences that resemble functional genes but seem to have no purpose. The presence of similar eta globin pseudogenes in humans and chimps has been used as an argument for common ancestry of the two species. The argument has two parts: that the pseudogene sequences actually have no function, and that God would not create similar non-functional sequences in humans and chimps. The latter argument is theological and resembles many other theological arguments that have been proposed and later abandoned. Theological arguments should not be relied on unless well supported by Scripture.*

*The argument that the eta globin pseudogene has no function is consistent with most of the data, although lack of function has not been demonstrated. Possible function is suggested by the location of the pseudogene and differences in the extent of divergence of “intronic” and “exonic” sequences. The possibility that the eta globin pseudogene provides a binding site for a molecule involved in gene regulation has not been ruled out. At present, the evidence from pseudogenes fits reasonably well into an evolutionary interpretation, for those who choose to make that interpretation. However, there is much about the operation of the genome in general, and pseudogene sequences in particular, that is not well understood. Rapid progress is being made in understanding how the genome operates, and it is reasonable to expect that greater understanding of the meaning of pseudogenes will be forthcoming.*

### INTRODUCTION

Theists and naturalists have long argued over whether nature provides evidence of design. Many theists have claimed that nature is so designed that one can infer the existence of a designer. Some theists have made

---

\*updated edition

the stronger claim that nature reveals a designer who is the Creator God revealed in the Bible. Many examples of apparent design have been described, ranging from the non-random properties of the universe to the intricate mechanism of a living cell. To the theist, these features speak clearly of the existence of an intelligent Creator who created with a purpose in mind.

Naturalists have responded with arguments of their own. One argument that is currently popular is the claim that many features in nature are not designed very well. It is affirmed that such poor design indicates either an inferior designer or no designer at all. Several examples of allegedly poor design have been proposed (e.g., Miller 1994). One of the most difficult examples for theists to explain is probably the existence of certain DNA sequences known as pseudogenes. This paper will explore some of the characteristics of pseudogenes and their relationship to the argument for or against design.

### **WHAT ARE PSEUDOGENES?**

Ordinary structural genes are made of DNA sequences that contain coded information for making a particular protein molecule. The information includes a start signal, coding for the sequence of amino acids needed to make up the protein, and a stop signal. Additional signals that regulate the timing of gene activity are found adjacent to the gene, and often also at some distance from it. The amino acid-coding sequence is often broken up into portions known as “exons,” which are separated by spacer sequences known as “introns.” Pseudogenes are DNA sequences that appear similar to functional genes, but contain important defects that appear to make them incapable of producing a functioning protein (Proudfoot 1980). Defects of pseudogenes may include lack of a start codon, presence of extra stop signals, and abnormal or absent flanking regulatory elements. It is thought that mutations in pseudogenes are neutral, and hence free from selection. The first report of a pseudogene was in 1977 (Jacq, Miller & Brownlee 1977). Since that time, a large number of pseudogenes have been described in humans and a wide variety of other species. The supposed defects of pseudogenes have been used as an argument that nature is too poorly designed to attribute its existence to special creation by a supernatural Designer (Miller 1994).

Two types of pseudogenes are known: unprocessed pseudogenes and processed pseudogenes. Processed pseudogenes are found on differ-

ent chromosomes from their functional counterparts. They are called “processed” because they appear to be altered copies of active genes. They lack introns (spacer sequences within a gene) and certain regulatory sequences located in front of the gene, they often terminate in a series of adenines, and are flanked by direct repeats. (“Direct repeats” are associated with movable genetic elements, which may in some cases play a role in inserting a pseudogene into a chromosome.) Processed pseudogenes may be complete copies of the coding sequence, or may be incomplete copies, or may have additional inserted sequences. They seem to be present only in mammals (Vanin 1985). Processed pseudogenes are believed to have arisen in a three step process. The first step is copying of the DNA message into an RNA transcript. The introns are then edited out of this transcript to produce a messenger RNA (mRNA) molecule. Finally, the mRNA is copied back into a chromosome in a process called reverse transcription (see Vanin 1985 for review; see Tchenio et al 1993 for an example). The L1 family of repetitive DNA sequences appears to be the result of this process (Jurka 1989).

Unprocessed pseudogenes are usually found within clusters of similar, functional sequences on the same chromosome (Harris et al. 1984). They typically have “introns” and flanking regulatory sequences resembling a functional gene. As with processed pseudogenes, expression of an unprocessed pseudogene is generally prevented by stop codons. Numerous other differences interpreted as deletions, insertions and point mutations may also be present. A truncated mRNA transcript may or may not be produced. Unprocessed pseudogenes are found in a wide variety of organisms. They are believed to have arisen by gene duplication, which produced an extra copy of the gene. The extra copy, not being needed, could accumulate mutations without harming the organism. Examples of unprocessed pseudogenes are present in the alpha-globin and beta-globin gene families (e.g., see Hardison & Miller 1993 and references therein).

### **THE ARGUMENT FROM SHARED MISTAKES**

When genes for equivalent proteins are compared in different species, they are often found to differ in sequence. In general, the more similar two species are taxonomically the more similar are their DNA sequences, both in general and for specific enzymes. Exceptions do occur, but the overall pattern is easily recognized. Two explanations

have been proposed for the observed pattern of similarities in molecular sequences.

One explanation for sequence similarities is that they are inherited from an evolutionary ancestor. Genes are similar because they are both inherited from a common ancestor. Sequence differences are attributed to accumulation of mutations since the species diverged from their common ancestor. A second, contrasting, explanation is that sequence similarities are due to common design for a similar function. Sequence differences may reflect functional differences, such as might be required for protein function in different metabolic environments, or regulatory function in different genetic backgrounds. It seems unlikely that sequence similarities could be due to chance, but some have been interpreted in this way (e.g., Djian & Green 1992).

Similarities in functional sequences for the same protein in different organisms are to be expected, since they perform similar functions; however, what about similarities in sequences, such as pseudogenes, that seem to have no function? Pseudogenes are commonly thought to be flawed copies of functional genes. It has been argued (Max 1987, Gilbert 1993, Miller 1994) that similar pseudogene sequences shared by two or more species are best explained as the result of common ancestry, assuming that an intelligent designer would not repeatedly make mistakes in creating genes. This can be called the “argument from shared mistakes.”

Comparison of DNA sequences from humans, chimp and other mammals reveals a considerable number of shared pseudogenes that are similar in sequence as well as in positional relationship to other genes. Humans and chimps have many similarities; this is interpreted as indicating a recent common ancestry for humans and chimps (Gilbert 1993). The best known example of a shared pseudogene is the eta globin (psi beta globin) gene, a member of the beta globin gene family.

### **THE BETA GLOBIN FAMILY AND THE (ETA GLOBIN) PSEUDOGENE IN HUMANS**

Human hemoglobin molecules are made of two sets of proteins, produced by the alpha globin genes and the beta globin genes. Both beta globin and alpha globin genes occur in “families” of non-identical copies. The beta globin gene family is located on the short arm of human chromosome 11 (11p15.5), near the gene for insulin (Lalley et al. 1989).

A family of alpha globin genes is also present in mammals, but it is located on a different chromosome (16p13).

The beta globin gene cluster consists of five somewhat similar functional genes and one pseudogene. The five functional genes are arranged on the chromosome in a sequence that corresponds to the sequence of timing of their respective functions during growth and development. The first gene in the series is the “epsilon globin” gene, which helps form hemoglobin molecules early in embryonic development. The second and third genes are called “gamma-G” and “gamma-A.” They help form hemoglobin molecules later during fetal development. The “eta globin” pseudogene is next in sequence, followed by the “delta” globin gene which is produced at a low rate in adults. The last gene in the series is the “beta” globin gene, which produces most of the adult beta globin, and gives the gene family its name. As the adult globin genes become functional, the fetal genes are turned off. The fact that the sequence of the genes on the chromosome matches the sequence of their activity in the developing organism seems unlikely to be the result of chance, and can easily be interpreted as the result of intelligent design.

The eta globin sequence has several characteristics of pseudogenes. It resembles the other members of the beta globin gene family, but is most similar to the gamma-A globin gene. However, it has some important differences. Compared with the gamma-A globin gene, the eta globin pseudogene lacks a start codon (AUG) in the appropriate position. It also has numerous extra stop codons which would be expected to prevent production of any protein. No mRNA transcript or protein product has been identified, and it appears that none is produced. No medical defect is known that is traceable to the loss of this pseudogene. In short, the eta globin sequence is not associated with any known function or defect, and appears to be incapable of producing a useful molecule.

The beta globin gene family is also found in other mammals. Sequences of the human gamma-A globin gene and eta globin pseudogenes from humans and several other species have been compared (Chang & Slightom 1984). The human gamma-A globin gene contains three exons (portions of the DNA that code for amino acids) of 92, 223 and 129 nucleotides, respectively, for a total of 444 nucleotides. The corresponding “exons” of the human eta globin pseudogene differ from the gamma-A globin gene exon sequences in 29, 38 and 43 nucleotide positions, respectively, for an overall difference of 24.8%. The gamma-

A globin gene has two introns, of 122 and 877 bases, respectively. These differ from the “intron” sequences of the eta globin pseudogene by 46-79% and 72-94%, respectively (my figures differ somewhat from those of Goodman et al. 1984, probably due to problems in aligning the sequences). The gamma-A-globin exons and pseudogene “exons” are more similar to each other than expected from random sequences, while the “intronic” sequences are so different that no relationship among them can be inferred.

### **COMPARISONS OF ETA GLOBIN PSEUDOGENES IN HUMANS AND OTHER PRIMATES**

The arrangement of the beta globin gene family in other primates is very similar to that in humans (Harris et al. 1984). Humans, chimpanzees and gorillas have the same number of beta globin genes arranged in the same sequence. In chimpanzees, the beta globin group is on chromosome 9, which is equivalent to human chromosome 11 (Lalley et al. 1989). Baboons have a similar arrangement, but the delta globin gene appears non-functional, and is classified as a pseudogene. The New World owl monkey has only one gamma globin gene, with a possible partial second gene (Meireles et al. 1995), but the arrangement of genes is otherwise the same as in humans. This is true also for the galago (“bush baby”; Hardison & Miller 1993). Among non-primates, the rabbit has only one gamma globin gene, but lacks the eta globin pseudogene, while the delta globin gene appears to be a pseudogene.

The DNA sequences of the eta globin pseudogene exons in humans, chimpanzees and gorillas are similar. The chimpanzee eta globin pseudogene exonic DNA differs from the human eta globin pseudogene at six nucleotide positions and from the corresponding gorilla pseudogene at seven positions. One of these differences The gorilla pseudogene exonic DNA has three differences from humans and seven from chimpanzees. This means that chimpanzee and gorilla eta globin exon sequences are both slightly more similar to the human pseudogene than to each other.

It is clear that the “exon” portions of the eta globin pseudogenes in humans, chimps and gorillas are highly similar. None of their differences involves any of the eight stop codons in the pseudogenes. Several potential initiation codons (AUG) are present, and one of the differences in the chimpanzee produces an additional potential initiation codon in the second exon. However, none of these is sufficient to support protein coding function.

## GENE DUPLICATION HYPOTHESIS

If evolution is to occur, new genes must somehow be produced. The most popular explanation for the evolution of new genes is that they are modified from extra copies of existing genes. This explanation is known as the gene duplication hypothesis (Ohno 1970). According to this hypothesis, functional genes may be duplicated accidentally. The duplicate gene is not needed by the organism. Both copies of the gene may be subject to selection until one of them suffers a disabling mutation, such as a premature stop signal. This disables the gene so it no longer has any function, and is no longer subject to natural selection. It has become a pseudogene, and all subsequent mutations are neutral. Over time, mutations accumulate in the pseudogene. Eventually, according to the theory, random mutations may produce a new gene with a new function (e.g., see Long & Langley 1993).

The gene duplication hypothesis, although widely accepted, is not without some theoretical and empirical difficulties. Assuming the original gene had been optimized by selection, mutations in the coding region of the duplicated gene prior to a disabling mutation would likely result in production of inferior protein molecules. Individuals with one gene that produced inferior protein products would likely be selected against. Spread of a duplicated gene should be difficult under these conditions.

This problem could be reduced if mutations destroyed the function of the extra gene copy early in its history. However, there are only three stop codons, while there are 61 codons for amino acids. One would expect mutations resulting in destruction of function to be much less common than those resulting in production of variant proteins, most of which could be expected to be inferior. Selection may also oppose maintenance of a pseudogene, since it may retain enough activity to disrupt normal cellular activities. Some pseudogenes are suspected to be involved in causing certain diseases (e.g., Wedell & Luthman 1993, Brakenhoff et al. 1994), which should result in negative selection against them. Thus, establishment and maintenance of a pseudogene by gene duplication may require a rather special sequence of events.

Walsh (1995) has calculated the theoretical conditions thought necessary for establishing the presence of a pseudogene in a population, assuming the pseudogene arose randomly by mutation. Establishment requires a high proportion of favorable mutations, a large number of reproducing individuals in the population, and a high selection coefficient.

It seems doubtful that these calculations can explain the frequency of pseudogenes in living species, and some other explanation would be preferred.

Another problem for the gene duplication hypothesis is that the existence of duplicate copies of a gene does not necessarily permit one of the copies to diverge from the others. For example, seven copies of the “Enhancer of split” gene are present in *Drosophila*, but it appears that none of them is free to mutate (Maier et al. 1993). The “duplicated copies” are not extra, but all seem to be required. Many genes occur in multiple copies that remain similar to each other rather than diverging. This has been explained as due to a process known as gene conversion, in which one DNA sequence is “converted” during copying to match another sequence. This may result in maintenance of similarity among several copies of a sequence. The situation in which multiple copies of a sequence maintain close sequence similarity is known as “concerted evolution” (e.g., Moore et al. 1993). Concerted evolution would tend to prevent divergence of duplicated genes, thus presenting a problem for the gene duplication hypothesis. Another problem with the gene duplication hypothesis is that tetraploid species have far fewer pseudogenes than would be expected (Larhammar & Risinger 1994).

Despite some difficulties in attributing evolution of new information from gene duplication, there seems to be evidence that gene duplication does occur. An apparent example of parallel gene duplications in flies has been described (in Menotti et al. 1991).

### **BETA GLOBIN GENES AND THE GENE DUPLICATION HYPOTHESIS**

It is thought that the eta globin pseudogene originated by duplication of a gamma-A globin gene, because of the similarity in their sequences. Both genes are present in all primates studied. Other mammals may have one or the other of the two genes. For example, gamma globin, but not eta globin, genes are present in rabbits; goats have eta globin but not gamma globin genes (Hardison & Miller 1993); the opossum has neither (Goodman et al. 1987).

It would be useful to review the evolutionary explanation for the distribution of eta globin genes in mammals. The proposed explanation is that the common ancestor of marsupials and placental mammals lacked both genes. After the evolutionary divergence of the marsupials, the gamma globin gene formed by duplication of an existing gene in the



beta globin family. Later, but before radiation of the orders of placental mammals, the eta globin gene formed from a duplicated gamma globin gene. This second supposed gene duplication is estimated to have occurred at least 140 million years ago (Harris et al. 1984). Gamma and eta genes must both have been present in ancestral placentals, but presumably gamma was lost by goats and eta was lost by rabbits.

According to this scenario, the eta gene must have been functional at first, because it is functional in goats. It is non-functional in all primates, which is interpreted to mean it was already nonfunctional in the ancestral primates. According to Martin (1993), primates probably originated in the Late Cretaceous, perhaps 70 to 80 million years ago. This interpretation implies that the eta globin pseudogene has been maintained for more than 70 million years without being converted into a useful new gene and without being eliminated. The persistence of a non-functional DNA sequence in an entire lineage for such a supposed long period of time seems remarkable in the context of the gene duplication hypothesis.

The gamma globin gene is believed to have duplicated a second time, producing the A-gamma and G-gamma genes. Humans, apes, Old World monkeys, and some New World monkeys have two functional gamma globin genes. Other mammals, including galagos, tarsiers and rabbits, have only a single gamma globin gene (Hayasaka et al. 1993, Hardison & Miller 1993). To explain this, the gamma globin gene is postulated to have undergone a second duplication after divergence of simians and tarsiers. Current interpretation of the fossil record of primates (Martin 1993) suggests that simians and tarsiers diverged during the Paleocene, perhaps 60 million years ago. It seems remarkable that both copies of a duplicated gene could remain functional for 60 million years if evolution has depended on gene duplication for the source of new genetic information.

### **THEOLOGICAL PRESUPPOSITION IN THE ARGUMENT FROM SHARED MISTAKES**

Several factors need to be considered in interpreting DNA sequence similarities in the eta globin pseudogenes. The argument has been presented that eta globin pseudogene similarities are compelling evidence of shared ancestry. This argument rests almost entirely on two assumptions: that the eta globin pseudogenes have no function; and that God would not create similar non-functioning sequences in separate species. Thus these assumptions must be carefully examined.

The argument that God would not act in a certain way is a theological argument, and can hardly be addressed by science. The validity of such an argument depends on the kind of God being postulated. The kind of God at issue for most of those involved in this discussion is the God who revealed Himself in the Bible. The question then is: What do the scriptures say about whether God would create structures or DNA sequences for which we can find no use in unrelated organisms? This subject is not addressed in the Bible, leaving us without an answer. We can postulate that God would not do such a thing, but this position would not be based on any evidence other than our own presuppositions, however reasonable they seem.

Another theological argument that has been advanced against some proposed actions of God is that God would not deceive us by acting in certain ways. This is equivalent to claiming that our understanding of nature can be trusted to accurately reveal God's activities. This argument is especially dangerous because it places human reason above divine revelation. The scriptures do state clearly that God does not deceive us, but they also make it clear that we are naturally prone to make wrong conclusions. The scriptures reveal the truth about history. When God tells us in scripture that He created in a certain way, we need not be deceived by what we believe to be appearances to the contrary. Our experience should teach us that much.

The argument that we can figure out what God would or would not do has not done well historically. At various times it has been claimed that God would create only perfectly circular orbits for the planets, or that God would create only perfect species that would not need to adapt to changing circumstances, or that God would not permit man to contaminate space. None of these arguments has survived. Claims about God's activities should be based on scripture.

### **SCIENTIFIC PRESUPPOSITION IN THE ARGUMENT FROM SHARED PSEUDOGENES**

A second assumption underlying the argument from shared mistakes is that shared pseudogenes, in this case the shared eta globin pseudogenes, have no function. Has it been demonstrated that these sequences have no function?

It is difficult to completely rule out any possibility of polypeptide production based simply on coding sequence. Examples are known in which the apparent DNA message is altered by RNA editing, reading

frame-shifting or skipping parts of sequences (Benhar & Engelbert-Kulka 1993, Dietz et al. 1993, Gesteland et al. 1992, Landweber & Gilbert 1993). Nevertheless, the available evidence seems to suggest that the eta globin pseudogene does not code for any protein. No RNA transcript or protein product has been identified. Each of the three “exons” contains at least one stop codon in each of the three “reading frames.” (“Reading frames” differ in which nucleotide of each base triplet is used as the starting point.) Seven potential start codons are present, but none of them is in “exon” one. These potential start codons are not sufficient for protein coding function. However, some pseudogenes may produce small amounts of polypeptides in specific tissues (Weinshank et al. 1991, Bristow et al. 1993, Misra-Ress, Cooke & Liebhaber 1994), so it is difficult to rule out the possibility that the eta globin sequence might produce a polypeptide.

DNA strands come in complementary pairs. One might wonder whether the DNA strand complementary to the pseudogene might have some function, but there seems to be no information available regarding this.

The eta globin pseudogene does not appear to function in chromosomal structure. Chromosomes are organized into loops that are attached at their bases to a nuclear material often called the nuclear scaffold. Scaffold associated regions are present within the beta gene cluster, and one of them is located near the eta globin pseudogene (Jarman & Higgs 1989). However, it appears that the scaffold associated region is not within the pseudogene sequence itself, making it unlikely that the pseudogene sequence functions in chromosomal structure.

The observation that the eta globin pseudogene is not associated with any known genetic defect is offered as further argument for its lack of function. Several hemoglobin beta globin abnormalities are known, but none of them is associated specifically with the eta globin pseudogene (Stamatoyannopoulos & Nienhuis 1994). This is interpreted as supporting the assertion that the pseudogene has no function. However, this argument is quite weak. The same result could occur for lethal mutations. No defective individuals would be observed because they do not survive long enough to be observed. Individuals with defective pseudogene sequences have been reported, but their abnormal hemoglobins were attributed to deleted portions outside the pseudogene sequence. It would be helpful to know whether normal individuals exist without the pseudogene sequence. Unless more information is available, the argument that

the eta globin pseudogene has no effect on health cannot be said to have been demonstrated.

The possibility that pseudogenes may have some function is worth exploring further. Some pseudogenes are believed to function as sources of information for producing genetic diversity (Fotaki & Iatrou 1993, Wedell & Luthman 1993), possibly involving a process similar to gene conversion. It is thought that partial pseudogene sequences are copied into functional genes, producing variants of the functional sequence. This phenomenon has been reported many times. Some examples include the immunoglobulins of mice (Selsing et al. 1982) and birds (Reynaud et al. 1989), mouse histone genes (Liu et al. 1987), and in horse globin genes (Flint et al. 1988) and human beta globin genes (Fullerton et al. 1994). The possible role of the eta globin pseudogene in gene conversion is unknown.

Regulation of globin genes is not fully understood, but several regulatory sites and protein factors have been identified (Stamatoyannopoulos & Nienhuis 1994). Each of the five functional beta globin genes has its own promoter region that participates in gene regulation. In addition, a locus control region (LCR) is found in a region several thousand bases upstream from the gene for epsilon globin, which is the first to be expressed.

There is no evidence that the eta globin pseudogene functions in gene regulation of the beta globin gene family (Engel 1993). However, that possibility has been suggested (Goodman et al. 1984, see also Vanin et al. 1980). The chromosomal arrangement of beta globin genes in a sequence corresponding to the timing of their activity is striking. It appears that chromosomal location plays an important role in beta globin gene regulation (Dillon et al. 1991).

The fact that the eta globin pseudogene is located between the fetal and adult genes suggests that it could play a role in gene switching—turning off the fetal gamma genes and turning on the adult beta gene. There is evidence that gene switching in human beta globin genes depends in some way on the sequence lying between the fetal and adult genes (Townes et al. 1991), although it is not known whether the eta globin sequence itself is involved. Some pseudogenes have been implicated in gene regulation (Singh & Brown 1991, Assinder et al. 1993, Koonin, Bork & Sander 1994). Such a role could involve competition for regulatory proteins, production of signal RNA molecules, or perhaps some other mechanism (e.g., see Enver et al. 1991).

Further suggestion of possible functionality of the eta globin pseudogene comes from a comparison of the “non-functional” sequences in humans and chimps. Non-functional sequences in this case include the A-gamma gene introns and the entire eta globin pseudogene. One would expect a similar rate of mutation in all non-functional sequences. We can test this by comparing the extent of difference between various regions of the non-functional sequences. Human and chimp A-gamma introns differ by 23 of 999 positions (2.3%). The respective eta globin “introns” differ by 16 of 999 positions (1.6%). The “exons” in the eta globin pseudogene differ by only 6 of 444 positions (1.35%). The figures for A-globin introns and eta globin exons differ by more than one-third. This could be explained as due to variations in the mutation rate, but this would tend to undermine the argument that differences in non-functional sequences are a function of time (the molecular clock hypothesis). It seems reasonable to suspect that mutations in the eta globin pseudogene “exons” are constrained, perhaps because it has some function that has yet to be discovered (cf discussion of *Drosophila* Adh locus in Sullivan et al. 1994).

Another presupposition of the argument from shared mistakes is that they could not have arisen independently, but must have been inherited from a common ancestor. Although convergence and parallelism are common problems in morphological studies (e.g., Carroll & Currie 1991), it seems improbable that identical nucleotide changes would occur independently. However, there is some evidence that nucleotide changes may not be random. Mutational “hotspots” (e.g., Hardison et al. 1991) have been identified, and independent gene duplication events have been inferred (Menotti, Starmer & Sullivan 1991).

### **ARE PSEUDOGENES “JUNK DNA”?**

It has been thought that only a small proportion of DNA codes for proteins. Typical estimates have been that perhaps 3% of the genome is involved. Recent discoveries (Wilson et al. 1994) indicate a figure closer to 30%. What is the function of the remaining portion? A large amount of DNA would be required for gene regulation, but this still leaves a significant part of the DNA with unknown function. That DNA fraction with no apparent function has been called “junk DNA.” Junk DNA has been thought to include intervening sequences (introns), satellite DNA (a highly repetitive DNA fraction), repetitive sequences, and pseudogenes.

As knowledge of the genome has increased, functions have been discovered for some of the sequences thought to be “junk” (Nowak 1994). For example, introns function in splicing together transcripts of exons. This constrains the kinds of changes that intronic sequences can tolerate. Some introns contain coding sequences which produce functional gene products (see Doolittle 1993 for review). Satellite DNA appears to be involved in chromosomal structure, especially at the ends (telomeres) and attachment points (centromeres) of the chromosomes. Repetitive DNA seems to have effects that are not well understood. Some diseases seem to be related to repetitive sequences (see Maddox 1994). It was recently noted that repetitive sequences seem to have a genomic arrangement characteristic of some kind of information code (Flam 1994), although the test used for this is apparently a weak test. Some supposed pseudogenes have been shown to be lowly or selectively transcribed (e.g., Yaswen et al. 1992, Imai et al. 1993, Vazeux, le Scanf & Fandeur 1993), which might suggest some function. The list of DNA sequences that have no effect on the organism has steadily decreased as knowledge of the operation of the genome has increased. This is reminiscent of the history of vestigial organs, in which apparent lack of function was actually lack of knowledge of what the function was. There is still much about pseudogenes that is not understood (Sullivan et al. 1994).

In retrospect, it seems perfectly reasonable to expect most DNA sequences, as well as organs, to have some function. One of the rules of nature seems to be that structures that are not useful tend to become lost. This is not to say that all DNA sequences must have a function. Copying errors, unequal crossing over and disruptive transposition all may contribute to the accumulation of useless DNA sequences. Many pseudogenes may indeed be junk DNA. However, the argument that particular DNA sequences must not have a function because we haven't discovered any function for them is an argument from silence. To conclude that pseudogenes are junk DNA seems premature.

### **SUMMARY AND CONCLUSION**

Pseudogenes are DNA sequences that resemble functional genes but seem to have no purpose. The presence of similar eta globin pseudogenes in humans and chimps has been used as an argument for common ancestry of the two species. The argument has two parts: that the pseudogene sequences actually have no function, and that God would not create

similar non-functional sequences in humans and chimps. The latter argument is theological, and is similar to many other theological arguments that have been proposed and later abandoned. Theological arguments should not be relied on unless well supported by scripture.

The argument that the eta globin pseudogene has no function is consistent with most of the data, although lack of function has not been demonstrated. Possible function is suggested by the location of the pseudogene and differences in the extent of divergence of “intronic” and “exonic” sequences. The possibility that the eta globin pseudogene provides a binding site for a molecule involved in gene regulation has not been ruled out. At present, the evidence from pseudogenes fits reasonably well into an evolutionary interpretation, for those who choose to make that interpretation. However, there is much about the operation of the genome in general, and pseudogene sequences in particular, that is not well understood. Rapid progress is being made in understanding how the genome operates, and it is reasonable to expect that greater understanding of the meaning of pseudogenes will be forthcoming.

#### ACKNOWLEDGMENTS

The author wishes to thank several reviewers for helpful comments and suggestions.

#### LITERATURE CITED

- Assinder SJ, De Marco P, Osborne DJ, Poh CL, Shaw LE, Winson MK, Williams PA. 1993. A comparison of the multiple alleles of *xylS* carried by TOL plasmids pWW53 and pDK1 and its implications for their evolutionary relationship. *Journal of General Microbiology* 139(3):557-568.
- Benhar I, Engelbert-Kulka H. 1993. Frameshifting in the expression of the *E. coli trpR* gene occurs by the bypassing of a segment of its coding sequence. *Cell* 72:121-130.
- Brakenhoff RH, Henskens HA, van Rossum MW, Lubsen NH, Schoenmakers JG. 1994. Activation of the gamma E-crystallin pseudogene in the human hereditary Coppock-like cataract. *Human Molecular Genetics* 3:279-283.
- Bristow J, Gitelman SE, Tee MK, Staels B, Miller WL. 1993. Abundant adrenal-specific transcription of the human P450c21A “pseudogene.” *Journal of Biological Chemistry* 268:12919-12924.
- Carroll RL, Currie PJ. 1991. The early radiation of diapsid reptiles. In: Schultze H-P, Truab L, editors. *Origins of the Higher Groups of Tetrapods*. Ithaca and London: Comstock Publishing Associates, p 354-424.
- Chang LYE, Slightom JL. 1984. Isolation and nucleotide sequence analysis of the beta-type globin pseudogene from human, gorilla and chimpanzee. *Journal of Molecular Biology* 180:767-784.
- Dietz HC, et al. 1993. The skipping of constitutive exons in vivo induced by nonsense mutations. *Science* 259:680-683.

- Dillon N, Fraser P, Hanscombe O, Greaves D, Lindenbaum M, Whyatt D, Strouboulis J, Hurst J, Grosveld F. 1991. Regulation of the human gamma-globin to beta-globin switch in transgenic mice. In: Stamatoyannopoulos G, Nienhuis AW, editors. *The Regulation of Hemoglobin Switching. Proceedings of the Seventh Conference on Hemoglobin Switching*, Airlie, Virginia, September 8-11, 1990. Baltimore and London: Johns Hopkins Press, p 34-44.
- Djian P, Green H. 1992. The involucrin gene of Old-World monkeys and other higher primates: synapomorphies and parallelisms resulting from the same gene-altering mechanism. *Molecular Biology and Evolution* 9:417-432.
- Doolittle RF. 1993. The comings and goings of homing endonucleases and mobile introns. *Proceedings of the National Academy of Sciences (USA)* 90:5379-5381.
- Drake JW. 1969. Comparative rates of spontaneous mutation. *Nature* 221:1132.
- Engel JD. 1993. Developmental regulation of human beta-globin gene transcription: a switch of loyalties? *Trends in Genetics* 9:304-309.
- Enver T, Raich N, Ebens A, Josephson B, Nakamoto B, Constantini F, Papayannou-poulou TH, Stamatoyannopoulos G. 1991. Autonomous and competitive mechanisms of human hemoglobin switching. In: Stamatoyannopoulos G, Nienhuis AW, editors. *The Regulation of Hemoglobin Switching. Proceedings of the Seventh Conference on Hemoglobin Switching*, Airlie, Virginia, September 8-11, 1990. Baltimore and London: Johns Hopkins Press, p 3-15.
- Flam F. 1994. Hints of a language in junk DNA. *Science* 266:1320.
- Flint J, Taylor AM, Clegg JB. 1988. Structure and evolution of the horse zeta globin locus. *Journal of Molecular Biology* 199:427-437.
- Fotaki ME, Iatrou K. 1993. Silk moth chorion pseudogenes: hallmarks of genomic evolution by sequence duplication and gene conversion. *Journal of Molecular Evolution* 37:211-220.
- Fullerton SM, Harding RM, Boyce AJ, Clegg JB. 1994. Molecular and population genetic analysis of allelic sequence diversity at the human beta-globin locus. *Proceedings of the National Academy of Sciences (USA)* 91:1805-1809.
- Gesteland RF, Weiss RB, Atkins JF. 1992. Recoding: reprogrammed genetic decoding. *Science* 257:1640-1641.
- Gilbert G. 1993. In search of Genesis and the pseudogene. *Spectrum* 22:10-21.
- Goodman M, Koop BF, Czelusniak J, Weiss ML. 1984. The eta-globin gene: its long evolutionary history in the beta-globin gene family of mammals. *Journal of Molecular Biology* 180:803-823.
- Goodman M, Czelusniak J, Koop BF, Tagle DA, Slightom JL. 1987. Globins: a case study in molecular phylogeny. *Cold Spring Harbor Symposia on Quantitative Biology* 52:875-890.
- Hardison R, Miller W. 1993. Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. *Molecular Biology and Evolution* 10:73-103.
- Hardison R, Krane D, Vandenbergh D, Cheng JF, Mansberger J, Taddie J, Schwartz S, Huang XQ, Miller W. 1991. Sequence and comparative analysis of the rabbit alpha-like globin gene cluster reveals a rapid mode of evolution in a G + C-rich region of mammalian genomes. *Journal of Molecular Biology* 222:233-249.
- Harris S, Barrie PA, Weiss ML, Jeffreys AJ. 1984. The primate psi beta 1 gene. *Journal of Molecular Biology* 180:785-801.
- Hayasaka K, Skinner CG, Goodman M, Slightom JL. 1993. The gamma-globin genes and their flanking sequences in primates: findings with nucleotide sequences of capuchin monkey and tarsier. *Genomics* 18:20-28.



- Imai K, Nakamura M, Yamada M, Asano A, Yokoyama S, Tsuji S, Ginns EI. 1993. A novel transcript from a pseudogene for human glucocerebrosidase in non-Gaucher disease cells. *Gene* 136:365-386.
- Jacq C, Miller JR, Brownlee GG. 1977. A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell* 12:109-120.
- Jarman AP, Higgs DR. 1989. Sites of attachment to the nuclear scaffold in the human alpha and beta globin gene complexes. In: Stamatoyannopoulos G, Nienhuis AW, editors. Hemoglobin Switching. Part B: Cellular and Molecular Mechanisms. NY: Alan R. Liss, Inc., p 33-45.
- Jeffreys AJ, Royle NJ, Wilson V, Wong Z. 1988. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* 332:278-281.
- Jurka J. 1989. Subfamily structure and evolution of the human L1 family of repetitive sequences. *Journal of Molecular Evolution* 29:496-503.
- Koonin EV, Bork P, Sander C. 1994. A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? *Nucleic Acids Research* 22:2166-2167.
- Lalley PA, Davison MT, Graves JAM, O'Brien SJ, Womack JE, Roderick TH, Creau-Goldberg N, Hillyard AL, Doolittle DP, Rogers JA. 1989. Report of the committee on comparative mapping. *Cytogenetics and Cell Genetics* 51:503-532.
- Landweber LF, Gilbert W. 1993. RNA editing as a source of genetic variation. *Nature* 363:178-182.
- Larhammar D, Risinger C. 1994. Why so few pseudogenes in tetraploid species? *Trends in Genetics* 10:418-419.
- Liu TJ, Liu L, Marzluff WF. 1987. Mouse histone H2A and H2B genes: four functional genes and a pseudogene undergoing gene conversion with a closely linked functional gene. *Nucleic Acids Research* 15:3023-3039.
- Long M, Langley CH. 1993. Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science* 260:91-95.
- Maddox J. 1994. Triplet repeat genes raise questions. *Nature* 368:685.
- Maier D, Marte BM, Schafer W, Yong Y, Preiss A. 1993. *Drosophila* evolution challenges postulated redundancy in the *E(spl)* gene complex. *Proceedings of the National Academy of Sciences (USA)* 90:5464-5468.
- Martin RD. 1993. Primate origins: plugging the gaps. *Nature* 363:223-233.
- Max E. 1987. Plagiarized errors and molecular genetics. *Creation/Evolution* 6(No XIX):34-45.
- Menotti RM, Starmer WT, Sullivan DT. 1991. Characterization of the structure and evolution of the Adh region of *Drosophila hydei*. *Genetics* 127:355-366.
- Meireles CMM, Schneider MPC, Sampiao MIA, Schneider H, Slightom JL, Chiu C-H, Neiswanger K, Gumucio DL, Czelusniak J, Goodman M. 1995. Fate of a redundant gamma-globin gene in the atelid clade of New World monkeys: implications concerning fetal globin gene expression. *Proceedings of the National Academy of Sciences (USA)* 92:2607-2611.
- Miller KR. 1994. Life's grand design. *Technology Review* (Jan/Feb) p 25-32.
- Misra-Press A, Cooke NE, Liebhaber SA. 1994. Complex alternative splicing partially inactivates the human chorionic somatomammotropin-like (hCS-L) gene. *Journal of Biological Chemistry* 269:23220-23229.
- Moore LA, Tidyman WE, Arrizubieta MJ, Bandman E. 1993. The evolutionary relationship of avian and mammalian myosin heavy-chain genes. *Journal of Molecular Evolution* 36:21-30.

- Morton BR, Clegg MT. 1993. A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbcl* in the grass family (Poaceae). *Current Genetics* 24:357-365.
- Neel JV. 1989. Human mutation rates. *Genome* 31:1104.
- Nowak R. 1994. Mining treasures from "junk DNA." *Science* 263:608-610.
- Ohno S. 1970. *Evolution by gene duplication*. NY: Springer-Verlag.
- Proudfoot N. 1980. Pseudogenes. *Nature* 286:840-841.
- Reynaud C-A, Dahan A, Anquez V, Weill J-C. 1989. Somatic hyperconversion diversifies the single  $V_H$  gene of the chicken with a high incidence in the D region. *Cell* 59:171-183.
- Rudd CJ, Daston DS, Caspary WJ. 1990. Spontaneous mutation rates in mammalian cells: effect of differential growth rates and phenotypic lag. *Genetics* 126:435-442.
- Selsing E, Miller J, Wilson R, Storb U. 1982. Evolution of mouse immunoglobulin lambda genes. *Proceedings of the National Academy of Sciences (USA)* 79:4681-4685.
- Singh M, Brown GG. 1991. Suppression of cytoplasmic male sterility by nuclear genes alters expression of a novel mitochondrial gene region. *Plant Cell* 3:1349-1362.
- Stamatoyannopoulos G, Nienhuis AW. 1994. Hemoglobin switching. In: Stamatoyannopoulos G, Nienhuis AW, Majerus PW, Varmus J, editors. *The Molecular Basis of Blood Diseases*. Philadelphia: W. B. Sanders Co., p 107-155.
- Sullivan DT, Starmer WT, Curtiss SW, Menotti RM, Yum J. 1994. Unusual molecular evolution of an *Adh* pseudogene in *Drosophila*. *Molecular Biology and Evolution* 11:443-458.
- Tchenio T, Segal-Bendirdjian E, Heidmann T. 1993. Generation of processed pseudogenes in murine cells. *EMBO Journal* 12:1487-1497.
- Townes TM, Ryan TM, Caterina JJ, Pawlik KM, Palmiter RD, Brinster RL, Behringer RR. 1991. Human globin gene regulation in transgenic mice. In: Stamatoyannopoulos G, Nienhuis AW, editors. *The Regulation of Hemoglobin Switching*. Baltimore: Johns Hopkins Press, p 16-33.
- Vanin EF, Goldberg GI, Tucker PW, Smithies O. 1980. A mouse alpha-globin-related pseudogene lacking intervening sequences. *Nature* 286:222-226.
- Vanin EF. 1985. Processed pseudogenes: characteristics and evolution. *Annual Review of Genetics* 19:253-272.
- Vazeux G, le Scanf C, Fandeur T. 1993. The RESA-2 gene of *Plasmodium falciparum* is transcribed in several independent isolates. *Infection and Immunity* 61:4469-4472.
- Walsh JB. 1995. How often do duplicated genes evolve new functions? *Genetics* 139:421-428.
- Wedell A, Luthman H. 1993. Steroid 21-hydroxylase (P450c21): a new allele and spread of mutations through the pseudogene. *Human Genetics* 91:236-240.
- Weinshank RL, Adham N, Macchi M, Olsen MA, Branchek TA, Hartig PR. 1991. Molecular cloning and characterization of a high affinity dopamine receptor (D1beta) and its pseudogene. *Journal of Biological Chemistry* 266:22427-22435.
- Wilson R + 55 authors. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* 368:32-38.
- Yaswen P, Smoll A, Hosoda J, Parry G, Stampfer MR. 1992. Protein product of a human intronless calmodulin-like gene shows tissue-specific expression and reduced abundance in transformed cells. *Cell Growth and Differentiation* 3:335-345.