Creation Research Society Quarterly 2013. 49:308–318.

The Origins and Genetic Functions of Pseudogenes

Jerry Bergman*

Abstract

Pseudogenes are genes that ostensibly lack the transcription or translation machinery required to produce protein. They are often used by Darwinists as evidence for the wasteful process that resulted from the evolution of the genome and, indirectly, as evidence for common descent because they appear to be evolutionary leftovers of past evolution life-forms. Recent research indicates that some or many pseudogenes do have a function, and several functions are discussed.

Introduction

Human genome mapping has identified enormous sections of DNA that do not code for proteins. This observation has been heralded as support for evolutionary naturalism (Wells, 2011). Called junk DNA (or non-informational DNA), it includes DNA sequences such as introns, repeated sequences, and pseudogenes (Bergman, 1994, 2001; Doolittle, 1993). Pseudogenes are a specific type of DNA; although very similar to protein coding genes, they are believed to be damaged genes and, as a result, no longer able to encode for protein (Chen et al., 2002). Although pseudogenes lack promoters and/or enhancers and appear to have crippling mutations such as frame shifts or premature stop codons, they "do tend to retain their characteristic intron-exon structure," which gives a hint of their function (Pink et al., 2011, p. 792).

Pseudogenes are an important topic in the creation-evolution controversy because neo-Darwinists commonly argue that a creator would not have created pseudogenes. Rather, they are the result of a blind, purposeless mechanism, such as evolution. Specifically, neo-Darwinists postulate that pseudogenes are evolutionary relics, molecular fossils that have accumulated during deep time and therefore can be used to help determine evolutionary history or phylogeny (Lee, 2003). This view interprets them as "long-dead genes ... that litter our chromosomes ... vestiges of old code associated with defunct [genetic] routines" (Gerstein and Zheng, 2006, p. 49).

History of Pseudogene Research

Pseudogenes were first reported in the literature by Jacq et al. (1977). In the years since this report, large numbers of pseudogenes have been identified in humans and a wide variety of other species. All animals, especially mammals, contain a large number of pseudogenes, but only a few putative pseudogenes have been identified in bacterial genomes (Andersson and Andersson, 2001, p. 829). The current estimate is that about 20,000 pseudogenes exist in humans alone, close to the number of functional genes (Svensson et al., 2006; Hirotsune et al., 2003). Because pseudogenes have sequences very similar to functional genes, they are often labeled "dead" or "disabled" genes (Hernandez et al., 1998). Normal genes contain DNA sequences that help to regulate or control the timing of various gene activities.

Pseudogenes appear to lack some or all of these controlling elements, such as a promoter site (RNA polymerase binding site for the start transcriptions), various regulators that control the level

* Jerry Bergman, Biology Department; Northwest State College; Archbold, Ohio, jbergman@northweststate.edu

Accepted for publication October 17, 2012

of protein produced, the specific code for the amino acid sequence of the protein produced during translation, and start and stop codons to control translation (Balakirev and Ayala, 2003; Gibson, 1994). Many pseudogenes also appear to contain extra stop codons (or premature stop signals) or have abnormal or missing flanking regulatory elements (Gibson, 1994).

It is theorized that these pseudogenes resulted from damage to the original genes by mutations, such as frame shifts, insertions and deletions. These mutations can result in gene nonfunctionality. They can also cause modifications that damage regulatory elements, especially initiation and termination sequences. As a result, the gene no longer produces functional proteins.

New research demonstrates that the regulatory DNA sequences are not always in the normal location close to the gene, or that they can always be readily identified. The fact that some of these signals are located a great distance from the gene presents a major difficulty in determining if a gene is a pseudogene. When these regulatory sequences are identified, technically the gene is not a pseudogene, even though it may still be identified as one. Consequently, we now recognize that there is not a clear distinction between genes and pseudogenes (Zheng and Gerstein, 2007).

Their similarity to functional genes has prompted Darwinists to consider pseudogenes evolutionary relatives of normally functioning genes (a reason for their extensive study). Richard Dawkins concludes that pseudogenes "are genes that once did something useful but have now been sidelined and are never transcribed or translated" (Dawkins, 2009, p. 332). In addition, he concluded that, "What pseudogenes are useful for is embarrassing creationists. It stretches even their creative ingenuity to make up a convincing reason why an intelligent designer should have created a pseudogene—a gene that does absolutely nothing and gives every appearance of being a superannuated version of a gene that used to do something" (Dawkins, 2009, p. 332).

Origin of Pseudogenes

Contrary to Dawkins' assumptions, pseudogenes are not contradictory to creation. The creation model accounts for genome degradation (Sanford, 2005). Thus, it anticipates the existence of some genes that have subsequently lost function since their original creation. This includes the pseudogenes that are categorized as disabled pseudogenes. These pseudogenes are complete genes that appear to be damaged by mutations that prevent transcription or translation. A classic example is the L-gulono-Y-lactone oxidase (GLO) gene in primates and humans that is missing several of the exons to code for a functional enzyme necessary for the manufacture of vitamin C (Cooper, 1999; Nishikimi et al., 1994).

However, not all pseudogenes can or need to be accounted for by degradation. One popular theory proposes that many pseudogenes arose by gene duplication from genetic mistakes, such as an unequal crossover, resulting in an extra copy of a functional gene, which is not immediately eliminated from the genome (see Bergman, 2006). If it did not have a detrimental effect on the organism, the duplicated gene could subsequently accumulate mutations. Mutations that occur in pseudogenes are assumed to be neutral (or near neutral). Thus, the organism would survive even though the mutation load in these genes is assumed to have slowly increased over time.

For this reason, mutations would be expected to accumulate in pseudogenes. The older the pseudogene, the more mutations it would contain. The gene duplication possibility for their origin has been suggested because most pseudogenes were not known to have a function and because pseudogenes are often similar to known functional genes. Gene duplication that results in useful genetic changes may have occurred, but this does not explain the origin of the original gene.

Types of Pseudogenes

Two basic classes of pseudogenes have so far been identified: *unprocessed* and *processed pseudogenes* (Wen et al., 2012).

Unprocessed

Unprocessed pseudogenes appear to be copies of active genes. The most common examples are alpha-globin and beta-globin pseudogenes. These unprocessed or duplicated pseudogenes are often found in clusters near similar, but functional, genes on the same chromosome (Brown et al., 1996).

This pseudogene type usually contains the same introns and flanking regulatory sequences as do functional genes (Doolittle, 1993). Expression of these pseudogenes is assumed to be prevented by the lack of transcription initiation sequences or premature stop codons that result in truncated proteins during translation. Another difference that also may exist between unprocessed pseudogenes and functional genes is that the former contains many putative point mutations, deletions, and insertions (Thiele et al., 2003).

Processed

Processed or retrotransposed pseudogenes may be another source of pseudogenes (Jackers et al., 1996). They are believed to result from a reverse transcription of processed mRNA, which results from retrotransposition followed by an insertion of the copy DNA somewhere else into the genome (Fairbanks and Maughan, 2006). Retrotransposons are products of transcription (i.e., RNA) that then serve as the template for production of DNA. These DNA sequences are then reintegrated into genomic DNA, thus resulting in a duplicated gene.

These so-called processed pseudogenes are commonly found on chromosome locations that are different from their functional counterparts.

Transcripts of DNA, called messenger RNA (mRNA), carry the instructions for the protein synthesis machinery in the cell. The mRNA sequence consists of exons, which code for specific proteins. These exons are separated on the RNA strand by noncoding spacer sequences called *introns*. Processing of mRNA involves removing the intron sequences and splicing the exons together to form an edited mRNA copy (Kandul and Noor, 2009). This processed copy is then used for the translation of proteins. Some pseudogenes have introns that are removed after transcription, suggesting that they once were functional genes. It also suggests that they are still functional, even if that function is currently unknown.

Processed pseudogenes are believed to be produced after the mRNA is edited to remove the introns. The mRNA is then copied back into DNA by a reverse transcription process. The copy of mRNA, called copy DNA or cDNA, is then integrated back into the chromosome. A specific non-pseudogene example is the L-1 family of repetitive DNA sequences (Jurka, 1989). Why this processing occurs is still unknown, but if past experience is any guide, the existence of this complex process indicates these pseudogenes have an important function.

Processed pseudogenes can produce complete copies of a specific coding sequence and may also contain additional inserted sequences. However, these pseudogenes lack certain other sequences necessary for protein translation. Because processed pseudogenes are produced from processed mRNA, they lack introns and the necessary upstream regulatory sequences (those existing in front of the gene) required for genes to direct and control protein assembly.

Conversely, exceptions exist, but these altered pseudogenes often terminate in a series of adenines (a poly-A-tail), as do functional genes. Most of these pseudogenes are also flanked by direct repeats: structures associated with movable genetic elements. These elements may play a role in assisting the process of splicing pseudogenes into chromosomes (Gibson, 1994).

Pseudogenes in the Origins Debate

Pseudogenes are of major interest in the origins debate. Although they closely resemble functional genes, they have no assigned, known function. Darwinists have assumed this means they have no function. Therefore, as Dawkins concludes, this is evidence for evolution and against intelligent design. Miller (1994, p. 25) argues:

> Intelligent design cannot explain the presence of non-functional pseudogenes unless it is willing to allow that the designer made serious errors, wasting millions of bases of DNA on a blueprint full of junk and scribbles. Evolution, in contrast, can easily explain them as nothing more than failed experiments in a random process of gene duplication in a genome of evolutionary remnants.

In other words, an intelligent designer would not create an organism with a large number of useless parts, and these worthless genes are evidence that the genome was not designed. Hence, a creator did not play a role in the origin of the genome.

Evolutionists also argue that the origin of pseudogenes is largely a result of chance, the outworking of natural law, the contingencies of history, and the deficiencies of the entire genetic system. This view, however, is changing as a result of continuing discoveries within molecular biology and genetics. Some commentators conclude the molecular revolution has produced more scientific knowledge in the past three decades than has accumulated in science's entire history, and if this trend continues, our view of the genome will no doubt continue to be revolutionized (Willingham and Gingeras, 2006).

In addition, some Darwinists also argue that beside duplicated genes, pseudogenes also provide an additional mechanism for evolutionary change. They argue that pseudogenes could allow for the existence of nonlethal mutations because the pseudogene no longer serves a genetic role in the organism. This permits the pseudogene to accumulate mutations that will not be deleterious to the organism. Thus, these mutations could evolve the pseudogene into a new functional gene that would confer a survival advantage to the organism. In other words, these Darwinists argue that mutated pseudogenes would *not* be lethal because the psuedogene is not necessary for life, but selection would favor it if enough mutations occurred so that the gene could become beneficial. In this way pseudogenes could evolve to become useful genes.

However, true pseudogenes cannot be selected for by natural selection because by definition they lack a selectable function. For the psuedogene to be selected, other mutations also must occur to cause the former psuedogene to be translated and obtain selectable activity (e.g., transcription/translation). If a gene can be translated, it is not a pseudogene and is assumed to have a function. A gene that cannot be transcribed will not be translated and will not undergo positive selection. So if a pseudogene evolved that was useful in the organism, it must also evolve the appropriate control sequences required for regulation and proper function. It could not undergo positive selection until all of the required proper control sequences first existed. If a new beneficial gene was able to acquire the control mechanisms to restore function, as evolutionists assume, so could a myriad of currently harmful genes.

Another problem with the evolutionary interpretation of pseudogenes is that many genes are pleiotropic (meaning the gene produces proteins that affect more than one function in the cell—see Bergman, 2010). Changing one gene will, consequently, affect other proteins and functions in the cell. In addition, most biological structures and biochemical systems, such as blood clotting, require many genes and thus are polytrophic, meaning all of the necessary parts must exist in a highly integrated functional set in order for the system to work (Behe, 1996). Thus, whether pleiotropic or polytrophic, single genes have interacting effects on the organism's phenotype. This interaction makes pseudogene evolution far more complicated than the mere reactivation by a few key mutations.

Pseudogene Patterns in Life

Researchers have found that unprocessed pseudogenes exist in a wide variety of organisms. When genes for equivalent proteins are compared in different species, the coding sequences often differ in both expected and unexpected ways. Likewise, the more taxonomically similar the two species are, often the more similar are their pseudogene DNA sequences. This is true in general and for specific enzymes, although some major exceptions exist in spite of this overall pattern. Another argument for their function is the fact that

> Pseudogenes are pervasive, and usually abundant, in all eukaryotic organisms … In human, about 12,000 DNA sequences show evidence of pseudogenes … comparative analysis of processed pseudogenes in the mouse and human genomes has surprisingly demonstrated that 60% of the processed pseudogenes are conserved in both mammalian species. The high abundance and conservation of the pseudogenes in a variety of species indicate that selective pressures preserve these genetic elements, and suggest that they may indeed perform important biological functions (Wen et al., 2012, p. 27).

Gene conservation is one of the strongest arguments for their function. If they were nonfunctional, mutations would accumulate, and they would not be conserved in the genome.

Comparisons of DNA sequences from primates, including humans and chimps, reveal a considerable number of pseudogenes that are very similar both in their sequence and positional relationship to other genes. The best-known example is the *eta-globin* gene, which is part of the β-like globin gene family. Eta-globin is classified as a pseudogene because it has no known start codon and several stop codons.

Evolutionists attempt to explain the sequence similarities of both genes and pseudogenes as a result of their inheritance from a common ancestor. The often-minor differences are attributed to the accumulation of mutations that have occurred since the species diverged from their hypothetical common ancestor. Conversely, creation explains the sequence similarities due to common designs required because they serve similar functions. Chimps, gorillas, and humans all have the exact same number of beta-globin genes, and they are arranged on the chromosome in the same order, a similarity that is a strong indication of functionality (Lalley et al., 1989).

Sometimes overlooked in the evolutionist's conclusions is the fact that the function of only a small number of proteins is fully understood, and sequence differences in many cases could reflect primarily differences required for the proteins to function properly in the various biological environments existing in different life-forms. The differences also could be the result of the regulatory needs of different animals. Likewise, pseudogenes would be logically similar if they have functions that constrain variation. The similarity of life is especially strong at the cellular level—the basic, undifferentiated eukaryotic cell found in all animal life is almost identical. Too much variation could cause malfunction, resulting in disease or even death.

Similarity Indicates Function

The fact that many pseudogenes are highly conserved (similar in both so-called primitive and advanced life-forms)—and many are very similar in humans and other organisms—supports the conclusion that most of them have a function. It also supports the conclusion that the genome is a complex, organized, designed system and not a haphazard accumulation of genes that originally formed as a result of mutations in an ancestral life form and then were inherited largely unchanged for eons down to our generation.

Similar pseudogene sequences in two related animals are considered to be orthologous, having originated from a common ancestor (Woodmorappe, 2000, Gibson, 1994, p. 91). This similarity in different animals is used by both Darwinists and creationists to argue that pseudogenes have a function. The fact that pseudogenes "exhibit evolutionary conservation of gene sequence, reduced nucleotide variability, excess synonymous over nonsynonymous nucleotide polymorphism, and other features that are expected in genes or DNA sequences that have functional roles" is evidence that they have one or more functional roles in the organism (Balakirev and Ayala, 2003, p. 123).

The question of pseudogene function is inextricably tied up with the role of all noncoding DNA, including introns, satellite DNA (highly repetitive short DNA sequences), and other repetitive sequences (Elder and Turner, 1995). Estimates of the amount of non-protein-coding DNA range from close to 5% for some bacteria, to 70% for some roundworms, to about 95% for humans (Gibson 1994, p. 104). Bacteria and vi-

ruses are said to have wall-to-wall genes. A significant fact is that, in general, the "higher" the life-form, the *greater* the percent of non-protein-coding DNA it has in its genome. The fact that more complex life-forms tend to have proportionally more noncoding DNA indicates that it has a function (Zheng and Gerstein, 2007), often being transcribed to regulatory microRNA (Djebali et al., 2012) or other regulatory features (Birney, 2012). The ENCODE project has documented several levels of function in many pseudogenes (Birney, 2012), including transcription, contradicting the once common conclusion that

> most pseudogenes lost transcribed activity either due to their integration into the silent region of the genome or due to mutation of the promoter or auxiliary regulatory elements. However, recent increasing evidence has demonstrated that pseudogenes represent a significant proportion of the "transcriptome" in various organisms … Evidence of expression of pseudogenes has been demonstrated not only in animals but also in plants (Wen et al., 2012, pp. 28–29).

Useless Gene Hypothesis

The useless gene hypothesis is critical for evolutionary assumptions about pseudogenes, but the theory tends to break down when the earlier putatively "less-evolved" organism stage is evaluated. For example, *Escherichia coli* sequence evaluations reveal wall-to-wall genes (4,253 genes and only 4.5 million base pairs of DNA). Humans have three billion base pairs and only about 23,000 genes (Mishra, 2010, p. 16). In bacteria, the ratio of protein-coding genes to base pairs of DNA is 1 to 1,050; in humans it is 1 to 100,000, a 95-times greater level (Zheng and Gerstein, 2007, p. 220).

If we were to assume that most pseudogenes are functionless, our understanding of genetics indicates that we would not expect that they would arise by the process of gene duplication gone awry. One reason this is evidence of deterioration, not evolutionary progress, is the fact that repetitive DNA (such as triplet repeat expansion disorders) is often associated with diseases such as Huntington's disease.

The function of many of the repetitive classes of non-protein-coding DNA is now understood. For example, one set of moderately repeated DNA codes is now known to code for transfer RNA and ribosome RNA (Zubay, 1995 p. 641). The genes that code for the five different histone types are usually clustered together on the genome, and this cluster forms a tandemly repeated array containing up to one hundred copies. Similarly, even globin and immunoglobulin genes are arranged as tandemly repeated arrays. These discoveries have reduced the amount of alleged junk DNA. In short, if

> pseudogenes are dysfunctional, why are they so highly expressed? Two possibilities may explain it. One possible explanation is that these pseudogenes are only incidental byproducts in the transcription events of other genes, because they are under the effect of the same promoters. An alternative explanation, which we are more inclined to accept, is that the pseudogene transcripts are in fact functional but not random products. More and more accumulating examples support this alternative explanation (Wen et al., 2012, p. 29).

Some Known Functions

As DNA research has progressed, functions for noncoding DNA have gradually been determined, and evidence that some or even most pseudogenes are functional has accumulated (Wen et al., 2012). Unprocessed pseudogenes are usually found on the same chromosome in clusters near similar but functional sequences, indicating both order and function. A major factor influencing this view is the discovery that, although the control system is generally just upstream of the coding exons, more and more exceptions are now known (Willingham and Gingeras, 2006).

Gene control is now recognized as enormously complex, and the distribution of regulation factors in the genome is so widespread that the gene concept, as a discrete region of DNA, has now been rejected (Pearson, 2006). As Balakirev and Ayala (2003, p. 123) conclude, "Pseudogenes that have been suitably investigated often exhibit functional roles, such as gene expression, gene regulation, and generation of genetic (antibody, antigenic, and other) diversity."

The observation that some genes that were once classified as pseudogenes have now been determined to have a function indicates that many more may be functional. An example is the chicken calmodulin gene, once considered nonfunctional but now known to have a function (Adams et al., 1992; Ohya and Yasuhiro, 1989). Some pseudogenes are transcribed but not translated, indicating a function for the transcript other than making protein. Studies of mouse pseudogenes indicate as many as 50% can be transcribed (Zheng and Gerstein, 2007). Tam et al., (2008, p. 534) found a subset of pseudogenes that generate endogenous, small interfering RNAs (endo-siRNAs), which

> are often processed from double-stranded RNAs formed by hybridization of spliced transcripts from protein-coding genes to antisense transcripts from homologous pseudogenes. An inverted repeat pseudogene can also generate abundant small RNAs directly. A second class of endo-siRNAs may enforce repression of mobile genetic elements, acting together with Piwi-interacting RNAs.

Furthermore, whole classes of small RNAs are used for regulation by blocking the translation of mRNA, and some may originate from pseudogenes (Guo et al., 2009).

Another possible potential function of pseudogenes is to bond to active genes during DNA replication to help stabilize the DNA. Some pseudogenes are involved in gene conversions or in recombination with functional genes to produce more mRNA transcript variations; a design used in order to create protein variations (Balakirev and Ayala, 2003). Willingham and Gingeras (2006, p. 1215) concluded the "widespread occurrence of noncoding RNAs—unannotated eukaryotic transcripts with reduced protein coding potential—suggests that they are functionally important." Some pseudogenes serve as "reservoirs of genetic variability," and this discovery has opened up a whole new area of research that may help to explain the enormous variety of life existing in the living world.

Pseudogenes are also postulated to have many other functions, including assisting active genes, gene silencing, and regulating development (Zheng and Gerstein, 2007). Much of this research is still ongoing, and the results are still very tentative. For example, evidence that pseudogenes may have a specific and important gene regulatory role has been derived from several studies. Further research, though, has questioned the findings of at least one study, that by Hirotsune et al., (2003) discussed below.

The Hirotsune Study

In 2003, evidence that a mouse pseudogene had regulatory role in development was reported. Hirotsune et al. (2003) obtained a line of transgenic mice that died soon after birth because of multiple organ failure. Subsequent investigations concluded that transcription of the *Mkrn1-ps1* pseudogene was disrupted in these mice by an insertion of the *sex-lethal* gene. In wild-type mice, *Mkrn1-ps1* pseudo-RNA was believed to be essential for the stability of a mRNA isoform produced from the parental gene makorin 1. The proposed function of *Mkrn1-ps1* was also supported by the finding that abnormal development of the transgenic mice could be reduced by *Mkrn1-ps1* or Mkrn1 by overexpression. The evidence was sufficiently compelling that the researchers concluded they had identified a functional mammalian pseudogene.

However, in a follow-up study, Gray et al. (2006) found evidence that *Mkrn1-ps1* was *not* transcribed and the transcript originally attributed to *Mkrn1-ps1* was an overlooked mRNA isoform at the *Mkrn1* locus. Furthermore, in the samples studied, it was found that the 5'-regions of both *Mkrn1-ps1* alleles were fully methylated, also indicating this pseudogene was likely not transcribed. These "contradictory results underscore the difficulty in evaluating the functional status of a pseudogene" (Zheng and Gerstein, 2007, p. 220).

Because a large number of proteins in a cell interact, either the interaction must be functional or at least the proteins cannot impede the roles of other proteins and cause adverse reactions. Consequently, a change in one protein often requires a change in a great number of other proteins. This is necessary to ensure that each protein's interactions with other proteins either has neutral or beneficial effects on the cell. Likewise, it now appears that some or many pseudogenes have a function that requires compatibility with the rest of the organism (Woodmorappe, 2000, 2003a, 2003b, 2003c). For this reason Zheng and Gerstein (2007) conclude that the boundary between pseudogenes and other genes is rather ambiguous.

Sometimes a functional mRNA is produced from what was assumed to be a pseudogene. This assumption is based up the presence of start codons that are either cryptic or located at a considerable distance from the gene's open reading frame. The extra stop codons in many pseudogenes may function as regulation elements that are selectively removed to allow for variation in the gene's expression. One study of a pseudogene found transcription begins after a stop codon, the new reading frame assumed to result from a reading frame formed by a frame shift mutation, which removed the stop codon. This new reading frame coded for a functional protein receptor (Balakirev and Ayala, 2003, p. 124).

Some pseudogenes could be functional genes that in humans were deactivated for reasons such as regulation or the result of a mutation. Thus,

> Pseudogenes may recover the full original function of the genes from which they derive. Pseudogene function has been restored in vitro by mutagenesis transfection, or in vivo by site-specific (or intermolecular) recombination (Balakirev and Ayala, 2003, p. 137).

When we learn more details of the role each gene has in a cell, or at least the function of the protein they encode, we will more readily ascertain what DNA sequences, if any, are in fact functionless. Only when the entire DNA genome is fully analyzed will it be possible to determine the percentage (if any) of DNA that is actually noncoding or lacks a function. As Balakirev and Ayala (2003, p. 137) write,

> How pervasive are "functional" pseudogenes? Many pseudogenes have been identified in all sorts of organisms on the grounds that they are duplicated genes that exhibit stop codons or other disabling mutations in their DNA sequences, so that they cannot have the full function of the original genes from which they derived. In many of these cases, however, it remains unknown, because it has not been investigated, whether the pseudogenes, described only on the basis of DNA sequences, may have acquired regulatory or other functions, or play a role in generating generic variability.

There are now indications that much DNA does not have a coding function, but most all of it has other roles in the

cell. Although the function of all human DNA is not yet known, it is clear that

> there is no evidence for significant stretches of DNA with no function. Frequently, genes with related functions are clustered. These clustered genes are usually transcribed into single expression units (messenger RNA's) containing the information for the synthesis of several functionally related proteins (Zubay, 1995, p. 642).

This same pattern of gene clustering also exists in the most "primitive" bacteria known (Zubay, 1995 p. 642. See also the ENCODE research, such as Djebali et al., 2012).

Because geneticists have not yet discovered a use for many pseudogenes does not prove that none exists. It simply shows that we do not yet know the function of many DNA transcripts. Behe concluded that evolutionists claiming pseudogenes have been proven to be useless are not able to cite scientific literature for support of this claim because evidence for this view "is nowhere to be found" (Behe, 1996, p. 26). As Mounsey et al. (2002, p. 770) noted, "proving that a gene is totally nonfunctional, and is therefore definitely a pseudogene, is impossible." Balakirev and Ayala (2003, p. 137) concluded that "some functionality [of pseudogenes] has been discovered in all cases, or nearly, whenever this possibility has been pursued with suitable investigations." Pink et al. (2011, p. 792) adds that, although

> some pseudogenes are transcriptionally silent, others are active, raising the question of whether their noncoding transcripts are a spurious use of cellular energy or instead harnessed by the cell to regulate coding genes. This question is particularly pertinent given the recent flurry of evidence suggesting that long noncoding RNAs play a critical role in regulating genomic function.

This observation may explain the fact that some putative pseudogenes are named after the functional gene they resemble. For example, the 4,846 bp pseudogene named the CEL-like gene (CELL) has a "striking homology" to the CEL gene, except the CELL pseudogene lacks exons 2–7 found in the CEL gene (Lidberg et al., 1992, p. 630). This putative pseudogene, as well as many other presumed pseudogenes, is transcripted. Thus, these examples challenge the original pseudogene definition that they are damaged genes. Rather, experimental verification suggests that such a definition seems to apply to only a "tiny proportion of the large number of pseudogenes present in a variety of genomes" (Zheng and Gerstein, 2007, p. 219). Some putative pseudogenes are linked together, which also suggests that they have a function (Lamerdin et al., 1996; Schutte et al., 2000; Savelyeva et al., 1998).

The fact that some pseudogenes are selectively transcribed indicates that they were misnamed and should be reclassified (Zheng and Gerstein, 2007). In short "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" (Gibson, 1994, p. 104; also see Djebali et al., 2012). This research is critical, for example, because of the evidence that pseudogenes are important in cancer causation, and it

> has been shown that pseudogenes are capable of regulating tumor suppressors and oncogenes by acting as microRNA decoys. The finding that pseudogenes are often deregulated during cancer progression warrants further investigation into the true extent of pseudogene function. In this review, we describe the ways in which pseudogenes exert their effect on coding genes and explore the role of pseudogenes in the increasingly complex web of noncoding RNA that contributes to normal cellular regulation (Pink et al., 2011, p. 792).

Pseudogenes as Evidence of Dysgenetics

Some of the differences found in both genes and pseudogenes can be explained as the accumulation of mutations that have occurred since Creation (Woodmorappe, 2004). The creationist view postulates that the original created organism was mutation-free, and its once-perfect genomic system has since degenerated, leading to some useless DNA sequences and numerous imperfect copies of genes. Mechanisms that could create pseudogenes include mutations, unequal crossing over that disrupts functional DNA sequences, and inappropriate transposition.

For this reason the existence of nonfunctional pseudogenes does not disprove design of the original genome. At most, it proves that some designed genes were corrupted as a result of mutations or mistakes in some genetic processes. Mistakes in gene copying commonly occur, and although the vast majority are corrected, some well-documented examples of non-corrected genes exist in the medical literature (Jorde et al., 1997). Up to two-fifths of all pregnancies now result in miscarriages, often because of DNA damage; a problem that indicates human DNA is now enormously corrupt compared to the original creation (e.g., see Meisenberg and Simmons, 2006, p. 153). Evidence that some pseudogenes are damaged genes also includes their involvement in disease.

Ultimately, some pseudogenes may no longer have a function due to genome deterioration. Evolution cannot explain how or why these pseudogenes (and the entire genome) originally evolved. In contrast, the Creation model explains that these genes were originally created with functions, but have subsequently lost these functions as genomic degeneration occurred following the Fall and the Curse (Genesis 3). In order for damaged genes, such as pseudogenes, to become functional (as proposed by some evolutionists), many complex proteins

that are compatible with the pseudogene are required to have evolved also. These include helicases and other proteins that pry apart the two DNA strands and align the transcription machinery at the proper location, and enzymes that stitch the nucleotides together into the appropriate polymer. Proteins are needed to strengthen the polymer (single-stranded binding proteins), and other proteins are required to insert the pseudogene copy back into the DNA (Adams et al., 1998).

A Creationist Perspective

The close similarity of putative damage in pseudogenes in different life-forms called "shared mistakes" is used to argue for evolution based on the analogy that identical typo errors or other mistakes in printed texts argue for plagiarism. Thus, identical putative errors in genes argues for a mistake in the postulated common ancestor that was passed down to its progeny (Woodmorappe, 2004). In support of this view, Dawkins (2009, p. 336) wrote that the "very existence of pseudogenes—useless, untranscribed genes that bear a marked resemblance to useful genes—is a perfect example of the way animals and plants have their history written all over them."

Neo-Darwinists argue that "God would not create similar non-functional sequences in humans" and other animals, such as chimps; thus, a common ancestor is the better explanation (Gibson, 1994, p. 91). Aside from the difficulty of judging what God would do, a mutation in a similar pseudogene may be in the same location in two animals for many reasons, such as it is actually functional or because a similar genetic mutation occurred due to the presence of a mutational *hot spot* (i.e., site of frequent mutations) at that location (Andersson and Andersson, 2001).

Many pseudogenes, such as those that lack promoters or contain evidence of a frame shift or nonsense mutations or loss of splice sites, do appear to be evidence of genetic degeneration. The latter explanation conforms to the Genesis Fall causing degeneration, not the Darwinian progression view. Deterioration is a prediction of the Creation model, while evolutionary advancement is a prediction of neo-Darwinism (Willaims, 1981, pp. 114–119). As Zheng and Gerstein (2007, p. 219) note, the discovery that some pseudogenes are functional raises questions:

> How should the concept of 'nonfunctional' be interpreted in defining pseudogenes? How could this finding be amalgamated with the established evolutionary theory, which often uses pseudogenes as nonfunctional and neutrally evolving DNAs for estimating various parameters in evolution. The scientific community, especially those dealing with molecular evolution and gene or pseudogene annotation, began to ponder these questions.

Possible functions for pseudogenes are reported almost monthly (Pink et al., 2011; Hernández, et al., 1998; Baertsch et al., 2008; Cooper and Kehrer-Sawatzki, 2008; Jegga and Aronow, 2008). The title of a recent article on this topic "Pseudogenes Are Not Pseudo Any More" is a good summary of the current state of the research (Wen et al., 2012, p, 27).

Conclusions

Arguing for evolutionary naturalism on the basis of DNA sequences that do not have a known function is an argument from ignorance (Balakirev and Ayala, 2003). Now that we know many pseudogenes are transcribed and have been determined to have a function, the most that can be concluded is that the function of many pseudogenes is currently unknown. As the history of the neo-Darwinism vestigial organ argument has documented, it was difficult to determine the uses of many structures until more knowledge about anatomy and physiology was gained (Bergman and Howe, 1990).

Until the entire human genome has been carefully and fully studied (a feat that may still be decades from completion), it is premature at this early stage of genome research to conclude that any gene or structure is completely without function (Zheng and Gerstein, 2007). Pseudogenes that have been carefully

> investigated often exhibit functional roles, such as gene expression, gene regulation, generation of genetic (antibody, antigenic, and other) diversity. Pseudogenes are involved in gene conversion or recombination with functional genes. Pseudogenes exhibit evolutionary conservation of gene sequence, reduced nucleotide variability, excess synonymous over nonsynonymous nucleotide polymorphism, and other features that are expected in genes or DNA sequences that have functional roles (Balakirev and Ayala, 2003, p. 123).

As a result, the following conclusion has now been supported by the evidence.

> Pseudogenes have long been labeled as "junk" DNA, failed copies of genes that arise during the evolution of genomes. However, recent results are challenging this moniker; indeed, some pseudogenes appear to harbor the potential to regulate their protein-coding cousins. Far from being silent relics, many pseudogenes are transcribed into RNA, some exhibiting a tissue-specific pattern of activation. Pseudogene transcripts can be processed into short interfering RNAs that regulate coding genes through the RNAi pathway (Pink et al., 2011, p. 792).

The latest conclusion is that the

> published evidence has shown that pseudogenes are not only transcribed, but also post-transcriptionally modulate their cognate genes by three distinct mechanisms: (1) natural antisense RNA suppression; (2) RNA interference by producing

> siRNAs; and (3) acting as decoys of stabilizing or disabling/inhibiting factors (Wen et al., 2012, p. 29).

Only when the various functions of each gene are fully understood can viable conclusions about uselessness be hypothesized. Even then, "proving that a gene unit is totally nonfunctional, and therefore definitely a pseudogene, is impossible" (Mounsey et al., 2002, p. 772). The discovery that some sequences labeled pseudogenes do perform genetic functions justifies the assumption that many, if not most, pseudogenes may eventually be found to have a function. Current research continues to support this conclusion. As Zheng and Gerstein (2007, p. 222) concluded, the "diverse ways that a genome sequence can realize its biological function have made it difficult to define a nonfunctional sequence."

The evolutionary view, involving mutational and natural selection, predicts a mechanism that facilitates, or at least allows, the evolution of the genome. This would involve transformation from one hypothetical pregene to the millions of different genes existing in the living world by known, plausible, naturalistic processes. This evolutionary idea would include a naturalistic means of producing new genes that are functional, and yet not detrimental, to allow some of these genes to confer a survival advantage to the organism.

As more and more animal and plant genomes are sequenced and analyzed, it is becoming increasingly clear that the design view fits the data much better. It is "clear that a whole genome is less like a static library of information than an active computer operating system for a living thing" (Gerstein and Zheng, 2006, p. 49). A design view predicts the entire genome was originally functional and that it is dominated by information, stability, design, purpose, and order, but has degraded somewhat over time (Williams, 1981). This complex system is evidence for order, design, and purpose for most pseudogenes, even in their present, post-Fall condition (Pitman, 2012). It is now recognized that

> the study of functional pseudogenes is just at the beginning. There remain many questions to be addressed, such as the regulatory elements controlling the cell or tissue specific expression of pseudogenes. But, definitely, the so-called pseudogenes are really functional, not to be considered any more as just "junk" or "fossil" DNA. Surely many functional pseudogenes and novel regulatory mechanisms remain to be discovered and explored in diverse organisms (Wen et al., 2012, p. 31).

The research so far has falsified Dawkins's boast that pseudogenes are useful only for embarrassing creationists. We do not need to stretch our "creative ingenuity to make up a convincing reason why an intelligent designer should have created a pseudogene," as Dawkins claims. We need only to understand the current scientific literature, and continue to increase our understanding of the genome and how it functions.

Acknowledgements

Clifford Lillo, Jody Allen, Jean Lightner, Mary Ann Stuart, and John UpChurch and several anonymous reviewers.

References

CRSQ: Creation Research Society Quarterly

Adam, E., M.R. Volkert, and M. Blot. 1998. Cytochrome c biogenesis is involved in the transposon Tn5-mediated bleomycin resistance and the associated fitness effect in Escherichia coli. *Molecular Microbiology* 28:15–24.

Adams, R., J. Knowler, and D. Leader. 1992. *The Biochemistry of the Nuclear Acids*. Chapman and Hall, New York, NY.

Andersson J., and S. Andersson. 2001. Pseudogenes, junk DNA, and the dynamics of Rickettsia Genomes. *Molecular Biology and Evolution* 18:829–839.

Baertsch, R., M. Diekhans, W. Kent, D. Haussler, and J. Brosius. 2008. Retrocopy contributions to the evolution of the human genome. *BMC Genomics* 9:1–19.

Balakirev, E.S., and F.J. Ayala. 2003. Pseudogenes: are they "junk" or functional DNA? *Annual Review of Genetics* 37:123–151.

Behe, M.J. 1996. *Darwin's Black Box*. Free Press, New York, NY.

Behe, M.J. 2003. A functional pseudogene? An open letter to nature. *Discovery*, May 13, 2003.

Bergman, J. 1994. Genetic function of introns as a means of reducing copying errors. *Ohio Journal of Science* 94:25.

Bergman, J. 2001. The functions of introns: from junk DNA to designed DNA. *Perspectives on Science and Christian Faith* 53:170–178.

Bergman, J. 2006. Does gene duplication provide the engine for evolution? *Journal of Creation* 20:99–104.

Bergman, J. 2010. The pleiotropy problem for evolution. *CRSQ* 46:284–289.

Bergman, J., and G. Howe. 1990. *"Vestigial Organs"are Fully Functional: A History and Evaluation of the Vestigial Organ Origins Concept*. Creation Research Society Books, Terre Haute, IN.

Birney, E. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74.

Brown, M.A., C.F. Xu, H. Nicolai, B. Griffiths, J.A. Chambers, D. Black, and E. Solomon. 1996. The 5' end of the BRCA1 gene lies within a duplicated region of human chromosome 17q21. *Oncogene* 12:2507–2513.

Chen, C., A.J. Gentles, J. Jurka, and S. Karlin. 2002. Genes, pseudogenes, and Alu sequence organization across human chromosomes 21 and 22. *Proceedings of the National Academy of Sciences* 99:2930–2935.

Cooper, S. 1999. *Human Gene Evolution*. Academic Press, New York, NY.

Cooper, D.N., and H. Kehrer-Sawatzki (editors). 2008. *Handbook of Human Molecular Evolution*. Volume 1 and 2. John Wiley & Sons, New York, NY.

Dawkins, R. 2009. *The Greatest Show on*

Earth: The Evidence for Evolution. The Free Press, New York, NY.

Doolittle, R.R. 1993. The comings and goings of homing endonucleases and mobile introns. *Proceedings of the National Academy of Science* 90:5379–5381.

Djebali, S., C.A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G.K. Marinov, J. Khatun, B.A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R.F. Abdelhamid, T. Alioto, I. Antoshechkin, M.T. Baer, N.S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M.J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O.J. Lou, E. Park, K. Persaud, J.B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, M. Schaffer, L.-H. See, A. Shahab, J. Skancke, A.M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S.E. Antonarakis, G. Hannon, M.C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, and T.R. Gingeras. 2012. Landscape of transcription in human cells. *Nature* 489:101–108.

Elder, J., and B.J. Turner. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *The Quarterly Review of Biology* 710:297–315.

Fairbanks, D.J., and P.J. Maughan. 2006. Evolution of the NANOG pseudogene family in the human and chimpanzee genomes. *BMC Evolutionary Biology* 6:12.

Gerstein, M., and D. Zheng. 2006. The real life of pseudogenes. *Scientific American* 295(2):49–55.

Gibson, L.J. 1994. Pseudogenes and origins. *Origins* 21:91–108.

Gray, T.A., A. Wilson, P.J. Fortin, and Robert D. Nicholls. 2006. The putatively functional *Mkrn1-p1* pseudogene is neither expressed nor imprinted, nor does it regulate its source gene trans. *Proceedings of the National Academy of Science* 103(32):12039–12044.

Guo, X., Z. Zhaolei, M. B. Gerstein, and D. Zheng. 2009. Small RNAs originated from pseudogenes: *cis*- or *trans*- acting? *PLoS Computational Biology* 5:1–15.

Hernández, J.M., B. Blat, C. Iruela, F. Vila, and J. Hernández-Yago. 1998. Identification of two processed pseudogenes of the human Tom20 gene. *Molecular & General Genetics* 258:117–122.

Hirotsune, S., N. Yoshida, A. Chen, L. Garrett, F. Suglyama, S. Takahashi, K. Yagami, A. Wynshaw-Boris, and A. Yoshiki. 2003. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 423:91–96.

Ivey, R.L. Jr. (editor). 2003. *Proceedings of the Fifth International Conference on Creationism*. *Technical Symposium Sessions*. Creation Science Fellowship, Inc, Pittsburgh, PA.

Jackers, P., N. Clausse, M. Fernandez, A. Berti, F. Princen, U. Wewer, M.E. Sobel, and V. Castronovo. 1996. Seventeen copies of the human 37 kDa laminin receptor precursor/p40 ribosome-associated protein gene are processed pseudogenes arisen from retropositional events. *Biochimica et Biophysica Acta* 1305:98–104.

Jacq, C., J.R. Miller, and G.G. Brownlee. 1977. A pseudogene structure in 5S DNA of Xenopus laevis. *Cell* 12:109–120.

Jegga, A.G., and B.J. Aronow. 2008. Evolutionarily conserved noncoding DNA. In Cooper, David N., and Hildegard Kehrer-Sawatzki (editors), *Handbook of Human Molecular Evolution*, Volume 1, pp. 853–860 John Wiley & Sons, New York, NY.

Jorde, L., J. Carey, and R. White. 1997. *Medical Genetics.* Mosby, St. Louis, MO.

Jurka, J. 1989. Subfamily structure and evolution of the human L1 family of repetitive sequences. *Journal of Molecular Evolution* 29:496–503.

Kandul, N.P., and M. Noor. 2009. Large introns in relation to alternative splicing and gene evolution: a case study of *Drosophila bruno*-3. *BMC Genetics* 10:1–16.

Lalley, P.A., M.T. Davison, J.A. Graves, S.J. O'Brien, J.E. Womack, T.H. Roderick, N. Creau-Goldberg, A.L. Hillyard, D.P. Doolittle, and J.A. Rogers. 1989. Report of the committee on comparative mapping. *Cytogenetics and Cell Genetics* 51:503–32.

Lamerdin, J.E., S.A. Stilwagen, M.H. Ramirez, L. Stubbs, and A.V. Carrano. 1996. Sequence analysis of the ERCC2 gene regions in human, mouse, and hamster reveals three linked genes. *Genomics* 34:399–409.

Lee, J.T. 2003. Molecular biology: complicity of gene and pseudogene. *Nature* 423:26–28.

Lidberg, U., J. Nilsson, K. Stromberg, G. Stenman, P. Sahlin, S. Enerback, and G. Bjursell. 1992. Genomic organization, sequence analysis, and chromosomal localization of the human Carboxyl Ester Lipase (CEL) Gene and a CEL-Like (CELL) gene. *Genomics* 13:630–640.

Meisenberg, G., and W. Simmons. 2006. *Principles of Medical Biochemistry*. Mosby, Philadelphia, PA.

Miller, K. 1994. Life's grand design. *Technology Review* 97(2):24–32.

Mishra, N.C. 2010. *Introduction to Proteomics: Principles and Applications*. Wiley, Hoboken, NJ.

Mounsey, A., P. Bauer, and I. Hope. 2002. Evidence suggesting that a fifth of annotated *Caenorhabditis* elegans genes may be pseudogenes. *Genome Research* 12:770–775.

Nishikimi, M., R. Fukuyama, S. Minoshima, N. Shimizu, and K. Yagi. 1994. Cloning and chromosomal mapping of the human nonfunctional gene for L-gulono-gamma-lactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man. *The Journal of Biological Chemistry* 269:13685–13688.

Ohya, Y., and A. Yasuhiro. 1989. Functional expression of chicken calmodulin in yeast. *Biochemical and Biophysical Research Communications* 158:541–547.

Pearson, H. 2006. Genetics: what is a gene? *Nature* 441:398–401.

Pink, C., K. Wicks, D. Caley, E. Punch, L. Jacobs, and D. Carter. 2011. Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA* 17:792–798.

Pitman, S.D. 2012. Pseudogenes and other forms of 'junk' DNA. http://www.detectingdesign.com/pseudogenes.html

Sakai, H., T. Itoh, and T. Gojobori. 2008. Processed pseudogenes and their functional resurrection in the human and mouse genomes. In Cooper, D.N., and H. Kehrer-Sawatzki (editors), *Handbook of Human Molecular Evolution* Volume 1, pp. 1203–1208. John Wiley & Sons, New York, NY.

Sanford, J. 2005. *Genetic Entropy & The Mystery of the Genome*. Ivan Press, Lima, NY.

Sasidharan, R., and M. Gerstein. 2008. Protein fossils live on as RNA. *Nature* 435:729–1001.

Savelyeva, L., A. Class, S. Gier, P. Schlag, L. Finke, J. Mangion, M.R. Stratton, and M. Schwab. 1998. An interstitial tandem duplication of 9p23–24 coexists with a mutation in the BRCA2 gene in the germ line of three brothers with breast cancer. *Cancer Research* 58:863–866.

Schutte, B.C., B.C. Bjork, K.B. Coppage, M.I. Malik, S.G. Gregory, D.J. Scott, L.M. Brentzell, Y. Watanabe, M.J. Dixon, and J.C. Murray. 2000. A preliminary gene map for the Van der Woude syndrome critical region derived from 900 kb of genomic sequence at 1q32-q41. *Genome Research* 10(1): 81–94.

Svensson Ö, L. Arvestad, and J. Lagergren. 2006. Genome-wide survey for biologically functional pseudogenes. *PLOS Computational Biology* 2(5):358–369.

Tam, O., A. Aravin, P. Stein, A. Girard, E. Murchison, S. Cheloufi, E. Hodges, M. Anger, R. Sachidanandam, R. Schultz, and G. Hannon. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453:534–538.

Thiele, H., M. Berger, A. Skalweit, and B. Thiele. 2003. Expression of the gene and processed pseudogenes encoding the human and rabbit translationally controlled tumor protein (TCTP). *European Journal of Biochemistry* 267:5473–5481.

Wells, J. 2011. *The Myth of Junk DNA*. Discovery Institute Press, Seattle, WA.

Wen, Y-Z., L-L. Zheng, L-H. Qu, F.J. Ayala, and Z-R. Lun. 2012. Pseudogenes are not pseudo any more. *RNA Biology* 9(1): 27–32.

Willingham, A., and T. Gingeras. 2006. TUF love for 'junk' DNA. *Cell* 125:1215–1220.

Williams, E. 1981. *A Creation Model for Natural Processes in Thermodynamics and the Development of Order*. Creation Research Society Books, Chino Valley, AZ.

Woodmorappe, J. 2000. Are pseudogenes 'shared mistakes' between primate genomes? *CEN Technical Journal* 14(3): 55–71.

Woodmorappe, J. 2003a. Pseudogene function: regulation of gene expression. *TJ* 17:47–52.

Woodmorappe, J. 2003b. Pseudogene function: more evidence. *TJ* 17:15–18.

Woodmorappe, J. 2003c. Unconventional gene behavior and its relationship to pseudogenes. In Ivey, R.L. (editor), *Proceedings of The Fifth International Conference on Creationism*, pp. 505–515. Creation Science Fellowship, Pittsburgh, PA.

Woodmorappe, J. 2004. Potentially decisive evidence against pseudogene 'Shared Mistakes.' *TJ* 18:63–69.

Zheng, D., and M.B. Gerstein. 2007. The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? *Trends in Genetics* 23:219–224.

Zubay, G. 1995. *Principles of Biochemistry*. Wm. C. Brown, Dubuque, IA.