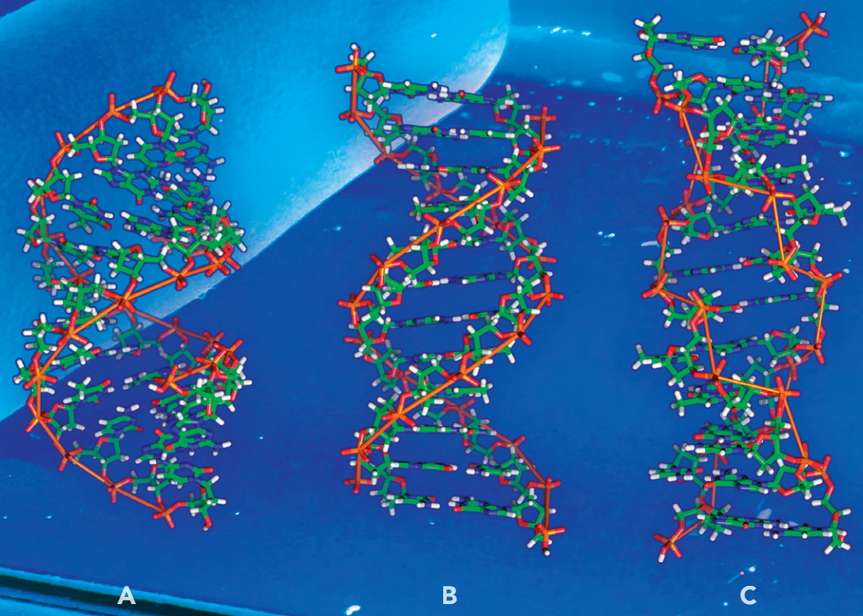
CREATION RESEARCH SOCIETY

QUARTERLY

Volume 52 Spring 2016 Number 4

SPECIAL ISSUE

Genetics



Creation Research Society Quarterly

Volume 52
Number 4
Spring 2016

Articles

Departments

Haec Credimus

For in six days the Lord made heaven and earth, the sea, and all that in them is, and rested on the seventh.—Exodus 20:11

Creation Research Society Quarterly

Volume 52
Number 4
Spring 2016

Cover by Michael E. Erkel, Afton, Virginia

Cover images:
Agarose gel with UV illumination - Ethidium bromide stained DNA glows orange. By University of Michigan School of Natural Resources DNA Lab, Ann Arbor (Creative Commons 2.0—https://creativecommons.org/licenses/by/2.0/).
The structures of A-, B- and Z-DNA. By Richard Wheeler (GNU Free Documentation License 1.2—https://www.gnu.org/licenses/#GPL) .
DNA sequence at the National Museum of Emerging Science and Innovation, Miraikan. By MIKI Yoshihito (Creative Commons 2.0—https://creativecommons.org/licenses/by/2.0/).

Table of Contents background photo: © Digoarpi | Dreamstime Stock Photos & Stock Free Images

The *Creation Research Society Quarterly* is published by the Creation Research Society, 6801 N. Highway 89, Chino Valley, AZ 86323, and it is indexed in the *Christian Periodical Index* and the *Zoological Record*.

Send papers on all subjects to the Editor: CRSQeditor@creationresearch.org or to Danny R. Faulkner, 1414 Bur Oak Ct, Hebron, KY 41048.

Send book reviews to the Book Review Editor: Don B. DeYoung, 200 Seminary Dr., Winona Lake, IN 46590, dbdeyoung@grace.edu.

All authors' opinions expressed in the *Quarterly* are not necessarily the opinions of the journal's editorial staff or the members of the Creation Research Society.

ISSN 0092-9166

Printed in the United States of America

The Challenge of Mount Improbable

A Special Issue of *CRSQ*

Kevin Anderson and Jean Lightner

More than 150 years ago, Charles Darwin proposed his ideas on the origin of species. He concluded that some traits could benefit organisms (e.g., make them faster) and some traits could hinder organisms (e.g., make them slower). Thus, those possessing beneficial traits had a greater chance of survival, and those with detrimental traits had a reduced chance of survival. The essence of this conclusion has become known as natural selection.

Darwin saw that variation in traits exists, but since his studies predated the field of genetics, he had no real understanding of the underlying basis for this variation. Differences in traits were taken as a given, and differences between species were generally attributed to the effects of natural selection. Even today some biologists claim that a particular trait arose by natural selection, as if traits magically appear when needed so natural selection can act upon them.

Not bound by any laws of heredity or an understanding of molecular genetics, Darwin saw "no reason to limit" the types of changes that an organism could undergo (Darwin, 1999, p. 127). With no limits, Darwin made some rather bold assumptions, not the least of which is that these changes could be so dramatic as to transform fish into amphibians or reptiles into mammals. He referred to this unlimited change as "descent with modification" (Darwin, 1999, p. 126). With this presumption, Darwin had outlined the basic concepts of universal common descent—the idea that all "animals and plants have descended from some one prototype," and all life shares a common ancestry (Darwin, 1999, p. 472).

About the same time that Darwin published *Origin of the Species*, an Austrian monk was conducting experiments with pea plants. From these experiments, Gregor Mendel observed that peas contain something he called *factors*, which caused the plant to grow tall or short and the pea pod to be yellow or green. These factors were also passed on to subsequent generations, affecting their growth and color as well.

Mendel presented his findings at a meeting of Austria's Brünn Society, but few listeners probably comprehended much of what they heard that day. Following his presentation, Mendel's work was published in the 1866 *Proceedings* of the Brünn Society. In an attempt to publicize his work, Mendel sent copies of his article to several well-known botanists and other naturalists. Interestingly, a copy of Mendel's article was reportedly found in Darwin's library. However, the copy was "uncut," indicating that Darwin had not even taken the time to slit the pages apart and read it (Henig, 2000). In fairness to Darwin, it would seem reasonable that, like most other naturalists of that time, even had he read Mendel's work, he would have had little comprehension of its significance. In fact, nearly fifty years would pass before the importance of Mendel's work would become recognized. During those same fifty years, though, Darwin's work would gain wider and wider acceptance (not coincidentally, without any genetic basis of the changes Darwin proposed).

What Mendel had unknowingly discovered was the inheritance patterns of chromosomal DNA. Mendel's *factors* were actually different versions of genes (known as alleles) carried on chromosomes. We now understand that changes to the chromosomal DNA nucleotide sequence (i.e., mutations) can alter the organism's physical features. Some mutations may cause the organism to grow a little taller or camouflage a little better. Other mutations may weaken the organism, such as reducing its line of vision or physical strength. Some mutations are lethal, resulting in death, while

still others may be "neutral," having no noticeable effect.

As Mendel's ideas became more broadly understood, evolutionists worked to incorporate this information into Darwinian teaching. In 1942, Julian Huxley introduced the term "modern synthesis" to reflect an updating of Darwinism to include basic Mendelian genetics (Huxley, 1942). Also known as neo-Darwinism, this updating was popularized by such noted evolutionists as George Gaylord Simpson and Ernst Mayr. However, even this updating predated an understanding of gene activity and molecular biology (a very significant absence).

Neo-Darwinism still represents the most popular form of evolution (almost exclusively what is presented in textbooks and journal publications). The most common version relies on random mutations to achieve the physical changes necessary for dramatic transformations; for example, amphibians evolving legs or birds evolving wings. However, random mutation alone creates an unwinnable game of chance. Even evolution apologist Richard Dawkins observes that "it is grindingly, creakingly, crashingly obvious that, if Darwinism were really a theory of chance, it couldn't work" (Dawkins, 1996, p. 77). Instead, Dawkins argues that natural selection sorts through the various traits, favoring those benefiting the organism and casting aside those that are detrimental. This combination is claimed to eliminate chance and achieve the needed changes.

Dawkins compares the attainment of biological complexity to that of climbing a mountain (i.e., climbing Mount Improbable). He recognizes the sheer difficulty of climbing this mountain and acknowledges that "only God would essay the mad task of leaping up the precipice in a single bound" (Dawkins, 1996, p. 77). He removes God from the process by envisioning the power of neo-Darwinism as its ability of

> breaking the improbability up into small, manageable parts, smearing out the luck needed, going around the back of Mount Improbable and crawling up the gentle slopes, inch by million-year inch. (Dawkins, 1996, p. 77)

As with most neo-Darwinists, Dawkins considers that the gradual accumulation of beneficial traits will slowly accomplish the task of originating enzymes, hormonal systems, respiratory organs, and brain activity. As such, according to Dawkins, random mutation combined with natural selection is equivalent to God.

A problem Dawkins fails to recognize is the sheer task of climbing Mount Improbable, even by small, incremental steps. Thirty years ago Michael Denton mused that "the credibility of [Darwinism] is weakened," not only by the high level of biological design known at that time, "but by the expectation of further as yet undreamt of depths of ingenuity and complexity" (Denton, 1986, p. 342). These expectations have certainly come to fruition. As a Berkeley biochemist laments, "It seems like we're climbing a mountain that keeps getting higher and higher…. The more we know, the more we realize there is to know" (Hayden, 2010, p. 664). Dawkins's Mount Improbable continues to get higher, steeper, and far more difficult to climb with each advancement of understanding. Since 1942, the growing knowledge of molecular genetics has not been kind to neo-Darwinism.

In this special issue of the *Creation Research Society Quarterly (CRSQ)*, Truman (2016) further highlights the unscalable nature of the mountain by offering an in-depth analysis of how living cells function as information processors. In this first of a two-part series, Truman describes the Boolean logic operations cells rely on, which involve multiple independent codes to accomplish the many and varied processes required to sustain life. Just as computer software does not design itself, neither do living cells design themselves.

Climbing Mount Improbable requires the generation of new beneficial mutations (Stoltzfus and Yampolsky, 2009), presumably resulting in new genes and genetic networks. Yet while beneficial mutations do occasionally appear, they are ineffective in climbing this mountain. In this issue of the *CRSQ*, Anderson (2016a) notes that certain popular examples of beneficial mutations cannot accomplish the needed transformations; new genes simply are not formed. Instead, the adaptive changes are the result of shuffling of preexisting genes, loss of gene expression, and loss of gene regulation. This loss is more appropriately compared to descending the mountain (Anderson, 2016b).

Interestingly, the absence of gene-forming mutations has led to the proposal that genetic loss is a driving force of Darwinian evolution. A "less-is-more" concept recognizes that most beneficial mutations are actually degenerative, not gene forming (Oh et al., 2015; Olson, 1999; Wang et al., 2006; Zhu et al., 2007). Organisms gain adaptive benefits from eliminating specific enzymes, regulatory systems, or transport proteins. However, universal common descent requires formation of new genetic systems. You cannot build specialized structures (such as feathers, legs, and wings) solely by eliminating preexisting genetic activity. Since the "less-is-more" concept cannot build the needed genetic systems, it assumes this is achieved by other, undocumented, mutations. Thus, the "less-is-more" model clearly is descending, not ascending Mount Improbable, but carries the assumption that the mountain must somehow have been scaled.

Molecular biologist James Shapiro states that randomness was originally inserted into Darwinism to exclude any hint of a creator. He adamantly insists that evolution must move beyond this original thinking to now include non-

random, directing genetic programs (Shapiro, 2011). Clearly, even mutations underlying adaptive change, which are accepted by creationists, often do not appear to be the result of random mutation. In other words, the neo-Darwinian mechanism of random mutation does not even account for the biblically recognized changes that, for example, allowed the canids on the ark to diversify into foxes in the desert, arctic, and other regions of the globe.

In this issue, Lightner (2016) summarizes some previous research in mammals suggesting that many mutations are not random and explores an enzyme used by the immune system to edit DNA in a highly regulated fashion. Again, while these mutations do not climb Mount Improbable, DNA editing enzymes may be able to explain why adaptive mutations are available when they are needed. This research lays a foundation for investigating whether this enzyme is at work in heritable germ-line mutations.

It is not just the assumption of random mutations that fails to support a climb up the mountain. Natural selection also has been shown to be an unsuitable component of this mountain-climbing venture. Various estimates illustrate it would likely take millions of years for just a few beneficial mutations to become fixed within a population (Durrett and Schmidt, 2007; Sanford et al., 2015). This is a major obstacle for Darwinism. There is simply not enough time for dramatic transformations, even using its extensive timescale. In this issue, three of the media reviews discuss the role of natural selection as it relates to the natural history of several different species and briefly hint at other factors important in adaptive changes since the Flood.

Several evolutionists have begun to agree with Shapiro (2011) and now recognize the need for a mechanism other than random mutation and natural selection. The *extended synthesis of evolution* is an attempt to integrate additional mechanisms (such as transgenerational epigenetics and multilevel selection) to compensate for the weaknesses of neo-Darwinism (Pigliucci and Müller, 2010). Others have a stated goal of seeking what they call the "third way"—an alternative to neo-Darwinism and creation (www.thethirdwayofevolution.com). By their own admission, they are still seeking.

Marshall (2015, p. 224) also concedes the need for a different mechanism, concluding that "if computer simulations have taught us anything, it's that gradual accidental 'Darwinian' processes never succeed in 'climbing Mount Improbable.'" Like Shapiro (2011), Marshall (2015) proposes that a directing program is part of an alternate version of evolution, which he labels evolution 2.0. This version entails specified alterations of DNA in response to certain environmental cues. As such, the mechanism of version 2.0 involves the action of "modular systems programmed to make sudden dramatic changes" (Marshall, 2015, p. 224).

If evolutionary changes are not blind and undirected, then where would such nonrandom genetic programming originate? Shapiro (2011) attempts to offer some scenarios, but ultimately fails, because his arguments require such programs to be an inevitable product of primordial blind and random processes. This is a problem every bit as significant for evolution to achieve as the problem these programs are intended to explain.

On the other hand, Marshall (2015) readily recognizes this problem and directly attributes the origin of these programmed modular systems to a creator. Interestingly, these modular systems are not unlike systems a biblical creation model also employs (e.g., hybridization, gene transfer, and epigenetics). With such created systems, organisms can present a wide variety of phenotypic traits.

However, Marshall still seeks to incorporate the unnecessary and genetically untenable baggage of universal common descent. He envisions that the action of multiple combinations of these modular systems will move an organism "from any one spot on the tree of life to any other" (Marshall, 2015, p. 144). Yet he is unable to support this claim with natural-occurring examples. No combination has ever been shown to move an organism from any one location on the tree of life to any other given location. This is strictly conjecture and very poor conjecture at that.

Using preexisting programs is not accurately an example of traversing Mount Improbable, since such a climb requires formation of new genes, new regulatory networks, and new genetic systems. Marshall's (2015) version 2.0 requires God to do all the "heavy climbing" but lacks any means of moving upward past this point on its own. There is no experimental data demonstrating his grand claim of moving anywhere along the tree of life (i.e., scaling the mountain's entire slope).

In addition, the bravado with which evolutionists claim significant evidence for common ancestry between humans and chimps is not in accord with the actual data. When comparing the DNA sequence between two or more species, certainly our imagination can always devise a story of how one species transforms into the other, regardless of their genetic relatedness. It is simply a matter of altering the DNA nucleotides as necessitated by the story.

This ignores the reality that there is far more to genetic activity than the nucleotide sequence. Chromosome function is not just a simple linear code (Riva, 2014). Specific genomic sequences are interweaved and multifunctional (Djebali et al., 2012). Interchromosomal interactions (i.e., chromosome kissing) facilitate three-dimensional topological domains between different chromosomes for specific gene activity

(Choudhury, et al. 2015). Gene-reading frames are buried within other reading frames, and many genes contain overlapping regions across both strands of DNA (Zhao et al., 2015). All proposed mechanisms of universal descent collapse in the face of this virtually insurmountable mountain.

What is more, changing the nucleotide arrangement can also disrupt proper cell function. Such disruptions can result in disease or even death of the organism. Thus, it is reasonable to suspect that there are many chromosomal regions where no path exists for the assumed DNA changes to occur. This is especially relevant when we consider that proper chromosome function is essential for a living organism.

Yet the assumption of common ancestry is so strong that it overrides all contradictory evidence. For example, Bradley (2008) claims there is excellent molecular support for a human-ape divergence 4 to 8 million years ago. Yet she acknowledges that this molecular evidence is often in direct conflict with morphology-based evidence, causing evolutionists to dramatically alter their account of how humans and primates are related. She further describes the decades it took for evolutionists to sort out the alleged divergence of humans, chimps, and gorillas because the different lines of data gave conflicting answers (Bradley, 2008). So physical similarities that were once promoted as excellent evidence of common descent are now explained as having arisen separately. Ways of explaining other aberrant data have also been suggested (Som, 2015). Thus, the assumption of common ancestry remains unquestioned, while its presumed evidence keeps changing radically.

Other claims of evidence supporting common ancestry between humans and chimps were put forth a number of years ago by popular evolution advocate Kenneth Miller (2007). He noted that chimpanzees, like other great apes, have 24 pairs of chromosomes, while humans have 23 pairs. Evolutionists assume that two chromosomes fused together as we humans traversed a separate evolutionary path than our chimpanzee cousins (Fan et al., 2002). Confident in his conjecture and premature conclusions about the clear evidence for the fusion, Miller willingly conceded, "If we don't find [evidence for the fusion], evolution is wrong. We don't share a common ancestor" (Miller, 2007). Contrary to Miller's confidence, further investigation makes it clear that there is no evidence for a fusion (Tomkins and Bergman, 2011; Tomkins, 2013); ergo, we do not share a common ancestor.

In this special *CRSQ* issue, Tomkins (2016) further investigates the human genome. He observes that humans have gene sequences that are remarkably different from that found in any other species where these genes appear. His in-depth look at these sequences not only highlights human uniqueness in relation to the animal kingdom but also shows how it directly contradicts universal common ancestry. As expected, considerable storytelling has been used by evolutionists in an attempt to accommodate the data, but the evidence is clearly more consistent with a biblical model, where humans were specifically created by God, and do not share a common evolutionary history with chimpanzees or any other organism.

In addition, genetic evidence supports a recent origin for humans. For example, human genealogies all overlap in the recent past. This indicates all humans descended from an ancestor that lived only about 3,000 BC (Rohde et al., 2004). Also, most of the single nucleotide variation affecting protein-coding genes among humans arose in just the last few thousand years (Fu et al., 2013).

Despite some uninformed criticism, there is also very good genetic evidence for the existence of an original human pair (i.e., Adam and Eve). In-depth analysis of this genetic data even shows it fits well in a few-thousand-year time frame (Jeanson, 2015). In this *CRSQ* issue, Carter and Lightner (2016) expand upon this concept. Exploring more of the DNA data, they describe genetic lineages that are consistent with the three daughters-in-law of Noah and the great population dispersion from Babel.

All of this genetic data is consistent with a recent origin of humans, lack of universal common descent, and overall failure of Darwinism to climb Mount Improbable. This mountain has not, and cannot, be climbed by mutations, transposition, or even limited preprogrammed systems. Evolutionists frequently offer examples of skids down the mountain (e.g., most mutations) or start at a higher point on the mountain by using modular systems (e.g., evolution 2.0), but none are able to account for scaling even small sectors of the mountain. Ascension of Mount Improbable requires the direct action of a creator. This occurred once (Genesis 1) and has not been repeated.

All the articles in this special issue provide a valid and dynamic description of the genetic basis of a creation model. Genetic data not known just a few years ago add to the failure of common descent and the vibrant explanatory power of biblical creation. There is clearly *not* a lack of genetic evidence for creation. Critics are generally either ignorant of this evidence or simply unwilling to grasp its significance.

References

Anderson, K. 2016a. Citrate utilizing mutants of *Escherichia coli*. *Creation Research Society Quarterly* 52:310—325.

Anderson, K. 2016b. How are new genes made? Answers in Genesis, April 21. https://answersingenesis.org/genetics/how-are-new-genes-made/

Bradley, B.J. 2008. Reconstructing phylogenies and phenotypes: a molecular view

of human evolution. *Journal of Anatomy* 212:337–353.

Carter, R., and J. Lightner. 2016. Human genetic data affirms biblical history on many levels and is an excellent resource for creation-based research. *Creation Research Society Quarterly* 52:249–255.

Choudhury, M., S. Zaman, J.C. Jiang, S.M. Jazwinski, and D. Bastia. 2015. Mechanism of regulation of 'chromosome kissing' induced by Fob1 and its physiological significance. *Genes & Development* 29:1188–1201.

Darwin, C. (1859) 1999. *The Origin of Species*. Reprint, Bantam Books, New York, NY.

Dawkins, R. 1996. *Climbing Mount Improbable*. W.W. Norton & Co., New York, NY.

Denton, M. 1986. *Evolution: A Theory in Crisis*. Adler & Adler Publishing, Bethesda, MD.

Djebali, S., and the ENCODE Group. 2012. Landscape of transcription in human cells. *Nature* 489:101–108.

Durrett, R., and D. Schmidt. 2007. Waiting for regulatory sequences to appear. *The Annals of Applied Probability* 17:1:1–32.

Fan, Y., E. Linardopoulou, C. Friedman, E. Williams, and B. Trask. 2002. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13–2q14.1 and paralogous regions on other human chromosomes. *Genome Research* 12:1651–1662.

Fu, W., T. O'Connor, G. Jun, H. Kang, G. Abecasis, S. Leal, S. Gabriel, M. Rieder, D. Altshuler, J. Shendure, D. Nickerson, M. Barnshad, NHLBI Exome Sequencing Project Members, and J. Akey. 2013. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493:216–220.

Hayden, E. 2010. Human genome at ten: life is complicated. *Nature* 464:664–667.

Henig, R.M. 2000. *The Monk in the Garden. The Lost and Found Genius of Gregor Mendel, the Father of Genetics*. Houghton Mifflin Co., New York, NY.

Huxley, J. 1942. *Evolution: The Modern Synthesis*. George Allen and Unwin, Ltd, London, UK.

Jeanson, N.T. 2015. A young-earth creation human mitochondrial DNA "clock": Whole mitochondrial genome mutation rate confirms D-loop results. *Answers Research Journal* 8:375–378.

Lightner, J. 2016. Adaptive genetic changes by design: a look at the DNA editing by activation-induced cytidine deaminase (AID). *Creation Research Society Quarterly* 52:265–274.

Kimura, M. 1979. Model of effective neutral mutations in which selective constraint is incorporated. *Proceedings of the National Academy of Sciences, USA*. 76:3440–3444.

Marshall, P. 2015. *Evolution 2.0: Breaking the Deadlock between Darwin and Design*. BenBella Books, Dallas, TX.

Miller, K. 2007. How to shut up pesky creationists. https://www.youtube.com/watch?v=dK3O6KYPmEw (accessed Jan. 18, 2016).

Oh, H., D. Choi, C. Goh, and Y. Hahn. 2015. Loss of gene function and evolution of human phenotypes. *BMB Reports* 48:7:373–379.

Olson, M.V. 1999. When less is more: gene loss as an engine of evolutionary change. *American Journal of Human Genetics* 64:18–23.

Pigliucci, M., and G.B. Müller. 2010. *Evolution—The Extended Synthesis*. The MIT Press, Cambridge, MA.

Riva, A. 2014. Information processing at the cellular level: beyond the central dogma. In Kasabov, N. (editor), *Springer Handbook of Bio-/Neuro-Informatics*, pp. 17–27. Springer Science & Business Media, New York, NY.

Rohde, D., S. Olson, and J. Chang. 2004. Modelling the recent common ancestry of all living humans. *Nature* 431:562–565.

Sanford, J., W. Brewer, F. Smith, and J. Baumgardner. 2015. The waiting time problem in a model hominin population. *Theoretical Biology and Medical Modelling* 12(1): 18.

Shapiro, J. 2011. *Evolution. A View from the 21st Century*. FT Press Science, Upper Saddle River, NJ.

Som, A. 2015. Causes, consequences and solutions of phylogenetic incongruence. *Briefings in Bioinformatics* 16:536–548.

Stoltzfus, A., and L. Yampolsky. 2009. Climbing Mount Probable: Mutations as a cause of nonrandomness in evolution. *Journal of Heredity* 100:637–647.

Tomkins, J. 2013. Alleged human chromosome 2 "fusion site" encodes an active DNA binding domain inside a complex and highly expressed gene – negating fusion. *Answers Research Journal* 6:367–375.

Tomkins, J. 2016. Human uniqueness and accelerated storytelling: how conserved regulatory regions in the genome challenge evolution. *Creation Research Society Quarterly* 52:256–264.

Tomkins, J., and J. Bergman. 2011. The chromosome 2 fusion model of human evolution—part 2. Re-analysis of the genomic data. *Journal of Creation* 25:111–127.

Truman, R. 2016. Cells as information processors, part 1. Formal software principles. *Creation Research Society Quarterly* 52:275–309.

Wang, X., W. Grus, and J. Zhang. 2006. Gene losses during human origins. *PLoS Biology* 4:e52.

Zhao, S., Y Zhang, W. Gordon, J. Quan, H. Xi, S. Du, and D. von Schack. 2015. Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. *BMC Genomics* 16:675.

Zhu, J., J. Sanborn, M. Diekhans, C. Lowe, T. Pringle, and D. Haussler. 2007. Comparative genomics search for losses of long-established genes on the human lineage. *PLoS Computational Biology* 3:e247.

Creation Research Society Quarterly 2016. 52:249–255.

Human Genetic Data Affirms Biblical History on Many Levels and Is an Excellent Resource for Creation-based Research

Robert W. Carter and Jean K. Lightner*

Abstract

Some have claimed that modern genetic data is at odds with biblical history. Yet closer examination reveals that the opposite is true. In terms of the origin of humanity, genetic data support the fact that all humans alive today can trace their ancestry back to a single male and a single female. When evolutionary assumptions are discarded and actual observable mutation rates are used, the molecular clock indicates that those individuals lived within a biblical time frame. Analysis of the human mitochondrial data reveals three major mitochondrial lineages, which appear to point to the three daughters-in-law of Noah. The Y chromosome distribution pattern supports a single paternally based dispersion as expected by the Babel event. Yet many questions remain, even as genetic data accumulate and computers make modeling more accessible to those outside the traditional university setting. The time is ripe for productive creationist research to answer important questions about the genetic history of humans using the wealth of data and tools now at our disposal.

Introduction

The history presented in Genesis makes it clear than humans were created in God's image, separately from all other animals (Genesis 1:20–27). Adam was created directly from the ground, and Eve was made from his side (Genesis 2:7, 21–22). As humans reproduced and filled the earth, the earth became filled with evil, so God chose to send a Flood to destroy the inhabitants (Genesis 6:5–7). Noah, his wife, his three sons, and their wives were the only humans that survived the global cataclysm (Genesis 6:18; 7:7, 13; 8:16; 1 Peter 3:20). All humans alive today have descended from them. Biblical data (Genesis 5, 11) and secular history enable us to estimate the time of Creation around 6,200 years ago and the Flood around 4,600 years ago (Hardy and Carter, 2014).

If this record is correct, it should be consistent with observations we can make today. Over the past several decades, an enormous amount of genomic data has been generated. This includes large-scale projects such as the HapMap

* Robert W. Carter, R.Carter@creation.info
Jean K. Lightner, jklightner@gmail.com
Accepted for publication February 10, 2016

project and the 1000 Genomes Project (International HapMap 3 Consortium, 2010; 1000 Genomes Project, 2015). While it is recognized that some errors are present in the data (Tomkins, 2011; Merchant et al., 2014; Carter, 2007), there should still be good agreement between the genomic data and the predictions one can make based on biblical history. Indeed, this has been affirmed in creationist journals (e.g., Carter, 2009; Jeanson, 2015), in Protestant theological journals (e.g., Sanford and Carter, 2014), and in a two-part article designed to reach out to Catholics (Sanford and Carter, 2015a, 2015b).

A major point of controversy involves the question of whether the currently observed human genetic variation is compatible with all humans descending from a single couple around 6,000 years ago. Dr. Francis Collins (a prominent evangelical Christian in the world of science, the former director of the Human Genome Project, and the current director of NIH) has gone on record as stating, "There is no way you can develop this level of variation between us from one or two ancestors" (Adkisson, 2011).

Similarly, Dennis Venema, Collins's fellow at the theistic-evolution promoting organization *Biologos*, has said:

> You would have to postulate that there's been this absolutely astronomical mutation rate that has produced all these new variants in an incredibly short period of time. Those types of mutation rates are just not possible. It would mutate us out of existence. (Haggerty, 2011)

Are these claims correct? How would we know and what, exactly, does the Bible predict about human genetics? This paper discusses some initial considerations essential to consistently interpreting the genetic data within a biblical framework. It will also lay some groundwork on what has been done, and what needs to be done, to model human genetic history from a biblical perspective. Such a model can help us understand our past (e.g., human migrations) and potentially may provide insights about human diversity as it relates to adaptation and disease.

Designed Diversity

All people on earth today have come about through the normal process of sexual reproduction. Gamete production in the mother and father created haploid versions of the parental genome through the process of meiosis. During this process, the complimentary copies of the parental autosomes recombined in large sections, gene conversion occurred, and mutations were introduced. Unlike everyone alive today, however, the genomes of Adam and Eve did not come about through natural processes. This is an incredibly important consideration for us and one that our opponents have rarely acknowledged. If Adam and Eve were specially created, we have multiple starting possibilities:

- Adam and Eve had unique genomes, with two original copies of each autosome (this is a good starting assumption); or
- Eve was a near clone of Adam, with the exception that she had no Y chromosome; or
- Eve was a haploid clone of Adam, essentially a product of meiosis, but with doubled chromosomes; or
- Adam and/or Eve were created with multiple genomes, possibly a different haploid set of chromosomes in each of their reproductive cells, essentially limiting future human genetic diversity only by the number of children they could potentially have.

Authors such as Collins and Venema are assuming there was no designed human diversity in Eden. According to that assumption, the four sets of chromosomes in Eden (two sets in Adam and two sets in Eve), would have all been identical. The only exception to this would have been the sex chromosomes (otherwise Adam and Eve would have necessarily both been female). This assumption is both unjustified and unreasonable. There is no reason to think any two chromosomes in Eden would have been identical. Even as Eden must have had designed sexual diversity (male and female), every chromosome could have carried unique alleles. Thus, the antediluvian population could have had much more genetic diversity than is seen among people today. Even if Eve was a near clone of Adam, Adam could have himself been heterozygous at tens of millions of nucleotide positions. Therefore, Venema's statement above is couched in error. He assumes he is starting from a blank slate, essentially a couple containing zero genetic diversity.

The available data can help us make estimates of created diversity in Adam and Eve. Theoretically, one of four nucleotide "letters" must appear at any position in the genome (A, C, G, or T). But when examining any specific location, one person might have a different letter in that position than another. Most variation is biallelic (in other words, only two letters are found at that location among all the people on earth), and there are millions of variable positions of this nature in the human genome (International HapMap 3 Consortium). Thus, any two people will have millions of single-letter differences among them. Yet, these variable locations are largely shared among all people groups, implying that this variation was established in the very early human population. From a biblical perspective, that means these variations had to predate the Babel dispersion, when the human population became fragmented linguistically and geographically (Genesis 11:1–9). Most reasonably, the majority of this genetic diversity would have been present in Adam and Eve at Creation, which could easily mean 10–100 million or more positions were created heterozygous (Carter, 2011).

During meiosis, homologous recombination shuffles the alleles (variants) between chromosomes. This occurs via crossing over and gene conversion. Since there are usually only one or two crossovers that occur per chromosome arm, large sections of DNA remain together on a chromosome as it is passed on. These regions are known as *haplotype blocks* and are recognized by a particular combination of alleles. Over many generations these regions should become more scrambled, shuffling the alleles, and resulting in haplotype blocks that are considerably smaller. Gabriel et al. (2002) estimate that most of the genome is contained in haplotype blocks of substantial size. The specific haplotypes and their boundaries were frequently shared across different populations of humans. All this is consistent with the population bottleneck at the Flood followed by a dispersion following the Babel incident several thousand years ago.

Mutation

The sequence of the genome can be changed by mutation. This could be a single nucleotide change, such as a transition from C to T. Alternatively, mutations may result in structural changes such as the duplication or deletion of a region, producing what is known as *copy number variants* (CNVs). It is now recognized that CNVs are very common sources of variation between humans. While some have no known effect, others are associated with adaptation or disease (Zarrei et al., 2015). Other structural rearrangements, such as inversions, can occur as well (Sudmant et al., 2015).

While there are even more alleles present in the human population that are attributable to mutation, most are not as widespread. Any individual human carries mostly common variants, which are likely created diversity, and fewer population-specific or even "private" alleles, which should largely be attributable to mutation (but see caveat under "The Effects of the Flood Bottleneck"). Detailed analysis of patterns in these alleles is important for understanding human genetic history as well as factors influencing adaptation and disease. Genetic variants that are widespread must have arisen early in human history; genetic variants that are very rare are much more likely to be "young" mutations. Interestingly, recent analyses by evolutionists have revealed that most protein-coding variants appear to be of very recent origin (Tennessen et al., 2012; Fu et al., 2013). Again, even though evolutionary assumptions were used in the estimates, the findings are consistent with the biblical historical parameters.

Historic Population Sizes

Speaking of the acceptable ranges of biblical parameters, historic population sizes are also important for us to consider. The size of a population dictates how much diversity it can hold, for small populations are subject to genetic drift: random sampling of the gene pool each generation can lead to significant changes in allele frequency in small populations. Genetic drift slows to a crawl in populations numbering in the thousands, and is essentially nonexistent in large populations.

Carter and Hardy (2015) used computer simulations to estimate the population sizes before the Flood, between the Flood and Babel, and within the nation of Israel during their sojourn in Egypt. The latter has been a frequent target of attack by skeptics who claim it is impossible for the Israelites to have attained the population size suggested in the Bible (Exodus 12:37–38; Numbers 1:46). On the contrary, simulations with some parameters indicate that attaining a population size of 2.7 million was possible within 215 years. If the Israelites were in Egypt longer, as many believe the Bible teaches, reaching such a population size was a trivial matter.

In contrast to the Exodus event, we do not have any biblical data that would allow us to estimate the population size at the Flood or at Babel. However, large population sizes at these events, and rapid reestablishment of large populations after each event, would have been relatively easy, given realistic population growth parameters. So, like the designed diversity example above, when we consider the relevant biblical parameters, there is no difficulty establishing appropriate population sizes in the given time. We are not limited to any particular population size, and thus the biblical model can handle data that demand either large or small historic population sizes. In other words, we have far more flexibility than many of our antagonists appear to assume.

The Effects of the Flood Bottleneck

Carter and Powell (2016) showed that the biblical claim that the entire human population was reduced to three reproducing couples is not problematic. There are multiple scenarios (assuming rapid population growth) in which almost no created diversity would be lost due to genetic drift. There are other scenarios (those with very slow growth, or if Noah's family were a small sample of the antediluvian population) where genetic drift would have been extreme. In high-drift scenarios, initial allele frequencies can rapidly change from 50:50 (the distribution they assumed in Adam and Eve) to extremely high/low allele frequencies. In these cases, a great deal of allelic fixation/extinction can occur, resulting in extensive loss of the initial allelic diversity. Intermediate levels of drift would result in partial loss of allelic diversity and a limited number of low frequency alleles (that are not derived by mutation).

When Carter and Powell (2016) compared their models to the real-world genetic diversity found among multiple

world populations, they concluded that modern humanity has experienced a large amount of genetic drift. This does not contradict the information in the paragraph above, but it does mean that of all the *possible* genetic history models, those with a strong bottleneck effect are more likely to reflect biblical human history. But when comparing Europeans to East Asians to Africans, they also saw that the allele frequency in one population was a strong predictor of the allele frequency in the other populations. In other words, the allele frequency spectrum was set up *prior* to Babel. Genetic drift must have occurred between Adam and Noah.

Mitochondrial DNA and Y chromosomes

Interestingly, it was the evolutionists who uncovered genetic evidence of a single woman (Mitochondrial Eve) and a single man (Y chromosome Adam) founding the human race. They also uncovered evidence of a severe population bottleneck, from which they construct their out-of-Africa model (Carter, 2010). These genetic situations are more consistent with a creation model than with evolution.

Mitochondria are organelles found in the cytoplasm of cells. They have some of their own DNA, which is separate from nuclear DNA yet considered part of the genome (all DNA of an organism). Mitochondrial DNA is passed down from mother to child, apparently with no contribution from the father. Based on differences in the sequence between people, it is clear we all could have come from one individual female, often called "Mitochondrial Eve."

Evolutionists place a time frame of when "Mitochondrial Eve" lived by assuming common ancestry between humans and chimps and the evolutionary timescale. However, when measured mutation rates in mitochondrial DNA were used, "Eve" was calculated to have lived around 6,000 years ago. Of course the evolutionists do not accept this time frame, so they have sought ways around the implications (Gibbons, 1998; Jazin et al., 1998). More recent in-depth analysis of mitochondrial DNA has upheld this biblical time frame for humans and found the same pattern in other organisms as well (Jeanson, 2014, 2015).

The out-of-Africa model was proposed by evolutionists to address the fact that patterns of genetic variability suggest a bottleneck occurred in the human lineage, and patterns of mitochondrial DNA variability across various populations suggested it may have originated from Africa (Cann et al., 1987). While various studies occasionally produce conflicting results, this is still the most popular evolutionary model of human history, partially because there is so much genetic diversity among Africans. The time frame and area from which humans dispersed differ from the Bible, but there are three major mitochondrial lineages that have been recognized (Witas and Zawicki, 2004; also see Figure 1). Carter (2009) has pointed out that there are other possible reasons for high genetic diversity in Africans, and (2010) that there is a more plausible ancestral sequence than the one proposed by evolutionists (Figure 2).

The human Y chromosome is remarkably similar among all humans, and the mutation rate is so slow it is difficult to detect (Jobling and Tyler-Smith, 2003). This is consistent with the biblical account, where Noah would have passed his Y chromosome on to his three sons less than 5,000 years ago. Yet, the chimpanzee Y chromosome is radically different from the human Y, which is a challenge for evolutionists to explain even in their extended time frame (Hughes et al., 2010). If humans and chimps had a common ancestor several million years ago, evolutionists are forced to propose that the Y chromosome mutated incredibly fast. But if all human males have very similar Y chromosomes (and they do), Y-chromosome Adam must have lived a very short time ago. Either way this is not consistent with evolutionary predictions. In contrast, this fits well with the biblical history of humans being created separately from all other animals.

Interestingly, global patterns in the Y chromosome suggest a less complex migration pattern than for mitochondrial DNA. It has been suggested that men generally have their families closer to their place of birth, and women leave their families to follow the men (Jobling and Tyler-Smith, 2003). This pattern is also consistent with the Babel dispersion, where families were spread according to identity of the fathers (Genesis 10:1–11:6), and so we would expect the mothers to be spread among the men.

Summary

The human genetic data is remarkably consistent with the biblical history. There is evidence that all humans trace their ancestry back to a single male and female, Adam and Eve. Genetic evidence points to a severe bottleneck, a dramatic decrease in population size, as we would expect from the Flood. Outside of Africa, there are three major lineages of mitochondrial DNA that would correspond to Noah's three daughters-in-law; yet there is a single worldwide lineage of Y chromosome that came from Noah through his three sons. Inside of Africa, the rarest sequences are also the most deviant. In other words, the out-of-Africa theory is based on statistical outliers! There is evidence of a single dispersion by families according to paternity, which corresponds well to the Babel event. When evolutionary assumptions are dropped and actual mutation rates are used, these events are within the biblical time frame.

Yet there is much information the Bible does not directly tells us, even while it does set limits for possible biblical models of human genetic history. For example, in Carter and Powell's

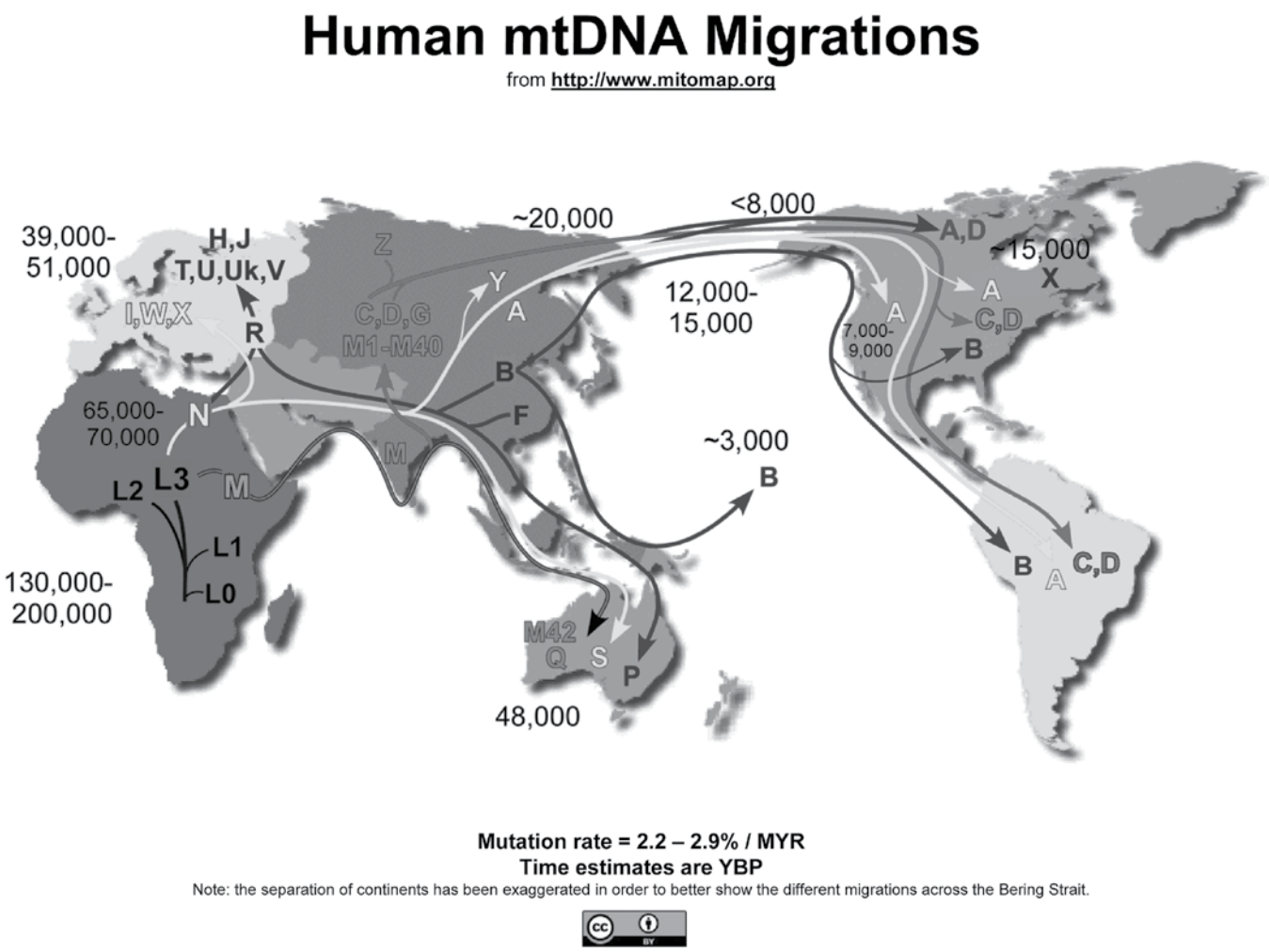


Figure 1. The evolutionary map of world migrations based on mitochondrial DNA has some striking similarities to predictions based on the biblical history. The out-of-Africa theory tells of a single dispersal of people, centered near and travelling through the Middle East, in the recent past. This type of pattern, with the migration originating in the Middle East, is predicted based on the history surrounding the Tower of Babel. Map from mitomap.org (http://www.mitomap.org/pub/MITOMAP/MitomapFigures/WorldMigrations2013.pdf).

model, the data forced them to conclude that either the antediluvian population was small or Noah and his wife and/or daughters-in-law were closely related. Is it unfair of us to appeal to a limited set of explanatory models when trying to fit the data to biblical history? Hardly, for this is exactly how the out-of-Africa theory developed (Carter, 2009), and it is still common practice among evolutionists today (Henn et al., 2016). Not only that, but most students of Creation and the Flood also have assumed the Flood bottleneck would involve a high degree of inbreeding, with possible loss of original diversity. This is especially true since Wieland's provocative 1994 article on the subject (Wieland, 1994). The inbreeding we might expect during the Flood/Babel period would produce exactly the allele frequency spectrum we see among modern people today.

Much of the discussion above could not have been part of any serious analysis of biblical history prior to just several years ago. The main reason for this has been the rise of powerful computers. With the rise of cloud computing, individuals now have inexpensive access to high-level computing resources once reserved for universities and governments. We would like to appeal to others interested in these subjects to build their own computer models. There are many questions remaining, and much refinement to existing conclusions can

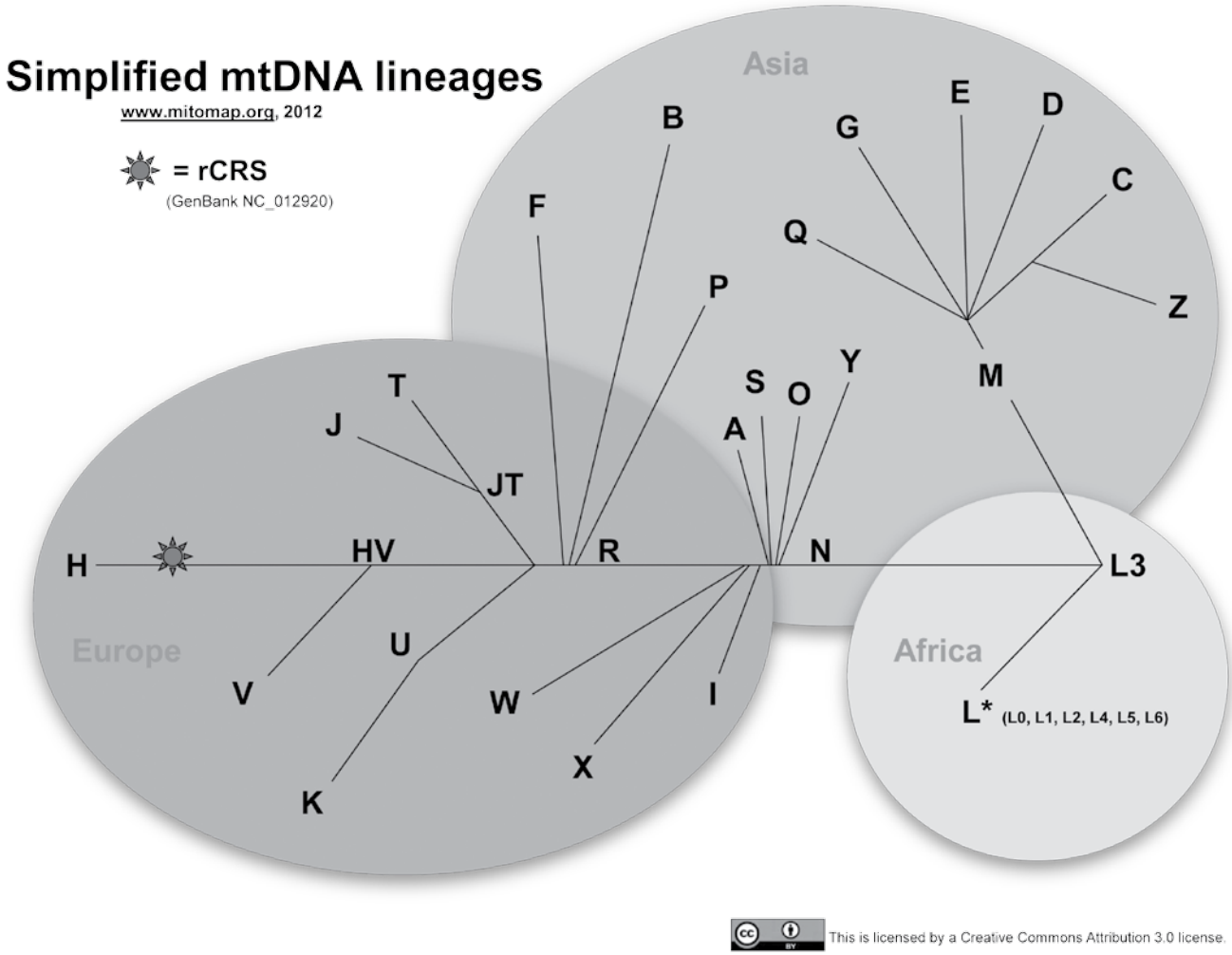


Figure 2. A diagram showing the relationship among major mitochondrial lineages. Evolutionists root the "tree" in Africa, but the sequences from there are statistical outliers. Carter (2007) placed the root at R, based on the most common nucleotides in each position across different human populations, but finding the location of the real root is a matter of statistics and historical uncertainties. Diagram from mitomap.org (http://www.mitomap.org/pub/MITOMAP/MitomapFigures/simple-tree-mitomap-2012.pdf).

be done. For example, if Neanderthals are human, how can we account for the presence of such genetically distinct humans that early in post-Flood history? And if Neanderthals interbred with humans early in modern human history (Kuhlwilm et al., 2016), what does this mean for the out-of-Africa theory since Neanderthals were supposedly not part of the bottleneck that led to the origin of "*Homo sapiens*?" And if sub-Saharan Africans came out of Babel, why do they display higher levels of genetic diversity than the rest of the world put together? These are fascinating questions, and as of right now they seem to be answered only by evolutionists. Creationists need to continue to develop competing robust models.

Robust creation models serve a purpose beyond just satisfying our curiosity about our history. A robust creation model that fits the data well can be used to make predictions, further test between the biblical history and the evolutionary one, and possibly give us valuable insights that relate to questions about adaptation and disease. These models would also help effectively counter challenges frequently leveled at biblical Creation. There is a tremendous opportunity for creation research in this area.

References

1000 Genomes Project. 2015. IGSR: The International Genome Sample Resource. www.1000genomes.org.

Adkisson, K. 2011. Noted scientist tackles

question of religious faith. *The Malibu Times*, June 29, 2011. www.malibutimes.com/news/article_3c135e3d-7695–5e22-b21c-9ceb8f752a7a.html (accessed February 13, 2016).

Cann, R.L., M. Stoneking, and A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.

Carter, R.W. 2007. Mitochondrial diversity within modern human populations. *Nucleic Acids Research* 35:3039–3045.

Carter, R. 2009. The neutral model of evolution and recent African origins. *Journal of Creation* 23(1):70–77.

Carter, R.W. 2010. Adam, Eve and Noah vs modern genetics. *Creation Ministries International*. www.creation.com/noah-and-genetics (accessed February 17, 2016).

Carter, R.W. 2011. The non-mythical Adam and Eve! Refuting errors by Francis Collins and *Biologos*. *Creation Ministries International*. www.creation.com/historical-adam-biologos (accessed February 15, 2016).

Carter, R., and C. Hardy. 2015. Modelling biblical human population growth. *Journal of Creation* 29(1):72–79.

Carter, R., and M. Powell. 2016. The genetic effects of the population bottleneck associated with the Genesis Flood. *Journal of Creation* 30(2):102–111.

Fu, W., T.D. O'Conner, G. Jun, et al. 2013. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493:216–220.

Gabriel, S.B., S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S.N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E.S. Lander, M.J. Daly, and D. Altshuler. 2002. The structure of haplotype blocks in the human genome. *Science* 296:2225–2229.

Gibbons, A. 1998. Calibrating the mitochondrial clock. *Science* 279:28–29.

Haggerty, B.B. 2011. Evangelicals question the existence of Adam and Eve. National Public Radio, 9 August 2011. www.npr.org/2011/08/09/138957812/evangelicals-question-the-existence-of-adam-and-eve?ft=1&f=1001.

Hardy, C., and R. Carter. 2014. The biblical minimum and maximum age of the earth. *Journal of Creation* 28(2):89–96.

Henn, B.M., L.R. Botigue, S. Peischi, I. Dupanloup, M. Lipatov, B.K. Maples, A.R. Martin, S. Musharoff, H. Cann, M.P. Snyder, L. Excoffier, J.M. Kidd, and C.D. Bustamante. 2016. Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proceedings of the National Academy of Sciences (USA)* 113(4):E440-E449.

Hughes, J.F., H. Skaletsky, T. Pyntikova, T.A. Graves, S.K.M. van Daalen, P.J. Minx, R.S. Fulton, S.D. McGrath, D.P. Locke, C. Friedman, B.J. Trask, E.R. Mardis, W.C. Warren, S. Repping, S. Rozen, R.K. Wilson, and D.C. Page. 2010. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* 463:536–539.

International HapMap 3 Consortium. 2010. Integrating common and rare genetic variation in diverse human populations. Nature 467(7311):52–58.

Jazin, E., H. Soodyall, P. Jalonen, E. Lindholm, M. Stoneking, and U. Gyllensten. 1998. Mitochondrial mutation rate revisited: hot spots and polymorphism. *Nature Genetics* 18(2):109–110.

Jeanson, N.T. 2014. New genetic-clock research challenges millions of years. *Acts & Facts* 43(4). www.icr.org/article/8017/ (accessed February 17, 2016).

Jeanson, N.T. 2015. A young-earth creation human mitochondrial DNA "clock": whole mitochondrial genome mutation rate confirms D-loop results. *Answers Research Journal* 8:375–378.

Jobling, M.A., and C. Tyler-Smith. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews: Genetics* 4:598–612.

Kuhlwilm, M., I. Gronau, M.J. Hubisz, et al. 2016. Ancient gene flow from early modern man into Eastern Neanderthals. *Nature* (early online release), doi:10.1038/nature16544.

Merchant, S., D.E. Wood, and S.L. Salzberg. 2014. Unexpected cross-species contamination in genome sequencing projects. *PeerJ* 2:e675; DOI: 10.7717/peerj.675

Sanford, J., and R. Carter. 2014. In light of genetics … Adam, Eve, and the Creation/Fall. *Christian Apologetics Journal* 12(2):51–97. media.wix.com/ugd/a704d4_f5f5f851384f4767889ad115b1e52f65.pdf.

Sanford, J.C., and R. Carter. 2015a. God, family, and genetics–a biblical perspective, part one: genetic evidences supporting the divine origin of man and family. *Kolbe Center*. kolbecenter.org/wp-content/uploads/2015/10/Rome Part 1 PDF.pdf (accessed February 17, 2016).

Sanford, J.C., and R. Carter. 2015b. God, family, and genetics–a biblical perspective, part two: genetic evidences refuting the evolution of man and family. *Kolbe Center*. kolbecenter.org/wp-content/uploads/2015/10/Rome Part 2 PDF.pdf (accessed February 17, 2016).

Sudmant, P.H., et al. 2015. An integrated map of structural variation in 2,504 human genomes. *Nature* 526:75–81.

Tennessen, J.A., A.W. Bigham, T.D. O'Conner, et al. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337:64–69.

Tomkins, J.P. 2011. How genomes are sequenced and why it matters: implications for studies in comparative genomics of humans and chimpanzees. *Answers Research Journal* 4:81–88.

Wieland, C. 1994. Decreased lifespans: have we been looking in the right place? *Journal of Creation* 8(2):138–141. www.creation.com/decreased-lifespans-have-we-been-looking-in-the-right-place.

Witas, H.W., and P. Zawicki. 2004. Mitochondrial DNA and human evolution: a review. *Anthropological Review* 67:97–110.

Zarrei, M., J.R. MacDonald, D. Merico, and S.W. Scherer. 2015. A copy number variation map of the human genome. *Nature Reviews: Genetics* 16:172–183.

Creation Research Society Quarterly 2016. 52:256–264.

Human Uniqueness and Accelerated Storytelling: How Conserved Regulatory Regions in the Genome Challenge Evolution

Jeffrey P. Tomkins*

Abstract

The Bible clearly states that humans were created in the image of God (Genesis 1:26–27). This makes us distinct in certain ways from the rest of the creatures God created, including primates. In addition to obvious outwardly visible trait differences, it would make sense that we would find certain regions of the genome that are distinctly different between humans and other animals, and this is in fact seen. Secularists postulate that these genetic differences arose from accelerated evolution since the time that humans allegedly diverged from apes; thus they call these regions *human accelerated regions* (HARs). HARs are exceedingly problematic for evolutionists due to the fact that they tend to be highly conserved across vertebrates but are markedly different in humans. However, within supposed vertebrate lineages, many of these regions are taxonomically isolated—they seem to arise suddenly—with no evolutionary history. A new phylogenetic analysis of 105 HAR genes in 10 different vertebrate taxa show that these sequences also display remarkable phylogenic discordance on a broad scale. This is inconsistent with the idea that these genes were generally conserved for tens or hundreds of millions of years but then suddenly evolved into taxonomically restricted forms. The data is more consistent with the creation model, wherein the genes that encode taxonomic distinction were custom designed.

Introduction

At the most fundamental level of objective discernment, even a child can clearly tell the difference between a human and a chimpanzee. However, the secular idea that humans somehow evolved from apes has become a leading icon of the evolutionary paradigm. In a creationist sense, this is one of the most objectionable components of the whole evolutionary paradigm because the Bible not only indicates that God made each creature "after its kind" but also that humans were uniquely created in God's image.

While many creatures exhibit distinct genetic differences, the issue of human relatedness to apes is seemingly bolstered in an evolutionary sense by

* Jeffrey P. Tomkins, The Institute for Creation Research, Dallas, TX, jtomkins@icr.org
Accepted for publication February 2, 2016

regions of high DNA similarity between humans and great apes (chimpanzees, gorillas, and orangutans), although pervasive inconsistencies, which evolutionists attribute to incomplete lineage sorting, negate a clear path of common ancestry (De et al., 2009; Ebersberger et al., 2007; Hobolth et al., 2007; Patterson et al., 2006). In addition, the DNA similarity paradigm, particularly in regard to human and chimpanzee DNA similarity, tends to be dominated by studies utilizing selective data that excludes genomic regions that are dissimilar (Bergman and Tomkins, 2012; Tomkins and Bergman, 2012).

Another problem in comparing human and ape DNA sequence is that great ape genomes, including chimpanzee, are computationally assembled from small individual sequence reads using the human genome as a reference sequence, and thus they appear to be more humanlike than they really are (Chimpanzee Genome Sequencing Consortium, 2005; Prado-Martinez et al., 2013; Tomkins, 2011). This problem is compounded even further by the fact that the chimpanzee genome is largely still a rough draft with numerous unsequenced gaps. In fact, a large number of studies, based on flow cytometry of nuclei and cytogenetic analyses of banding patterns, estimate that on average the chimpanzee genome is about 8% larger than human with significantly more heterochromatic DNA (Formenti et al., 1983; Koop et al., 1986; Pellicciari et al., 1982; Pellicciari et al., 1988; Pellicciari et al., 1990a; Pellicciari et al., 1990b; Seuanez et al., 1977). At present, it appears the alignable regions of the human and chimpanzee genomes are on average about 88% similar (Tomkins, 2015b). Nevertheless, there are many regions of apparent similarity between the genomes that are about 98% identical. It is these regions that are typically compared by evolutionists because they are conducive to hypothetical analyses regarding selection.

One of the features of the human genome that has been of particular interest to evolutionists during the past decade is termed human accelerated regions (HARs). These regions are a double-edged sword for the evolutionary paradigm in that they are both highly conserved (similar across taxa) yet markedly different in humans compared to other animals (particularly chimpanzees). Therefore, there is interest in finding such sequences and functionally characterizing them, as such sequences may help us understand what makes us uniquely human.

The detection of alleged accelerated regions of evolution assumes that evolution on a grand scale has actually occurred and requires a significant amount of hypothetical modeling. Under this assumption, DNA substitution rates are estimated based upon highly similar genomic regions from humans, great apes, and other vertebrates. These regions are so similar that they generally do not contain many sequence gaps (insertions or deletions) between taxa. In other words, the differences are primarily in single bases, called substitutions.

Early Discoveries of HAR and the Enigma of HAR1

The first popularized discovery of a HAR (demarcated HAR1) was a 118 base pair (bp) region that showed 18 base substitutions compared to its counterpart in the chimpanzee genome (Pollard et al., 2006a). When this region was assessed for variability among humans, it was found to be fixed in human populations (nonvariable). Making this discovery even more remarkable was that when the same genomic segment from chimpanzee and chicken were compared, there was only a 2 base difference out of the 118 bases. In the evolutionary mindset, the region clearly was highly conserved across taxa, but why was it so different between humans and chimpanzees? Hence the name human accelerated region is based on the evolutionary belief that it must have changed very rapidly after humans diverged from chimps. While scientists found the data to be especially intriguing, the results defied the evolutionary paradigm of slow and gradual evolution of the genome.

Even more intriguing was the fact that homologs for HAR1 could not be found in frog or any fish genome (Pollard et al., 2006a). Therefore, since it was functional and present in a chicken-like common ancestor (presumably about 310 million years ago according to evolutionary theory), then it originated "suddenly" on the evolutionary scene in some vertebrate ancestor about 400 million years ago. In this light, the HAR1 sequence appeared suddenly in vertebrates with no evolutionary precursor.

However, the homology mystery does not stop here. As it turns out, the HAR1 region is a part of an overlapping set of genes called HAR1A and HAR1B (previously referred to as HAR1F and HAR1R). Whereas the HAR1 region itself is highly conserved across taxa, even in chickens, the larger gene region of which it is only a small part is highly *nonconserved* and is very different between vertebrate taxa. So how could one small isolated segment in this region stay relatively the same during millions of years of evolution, while the surrounding region that it is intimately connected with changed so markedly? In fact, even in rhesus (a monkey) most of the entire HAR1A/B gene region of approximately 9,000 bases is almost completely unalignable to human (Pollard et al., 2006a).

The HAR1A and HAR1B genes produce noncoding RNAs and are expressed in the developing neocortex (Pollard et al., 2006b). As it turns out, the 18-base difference between the human and chimpanzee versions of the HAR1 gene lead to remarkably distinctive secondary structures, as shown in Figure 1 and described in detail as a result of a thorough

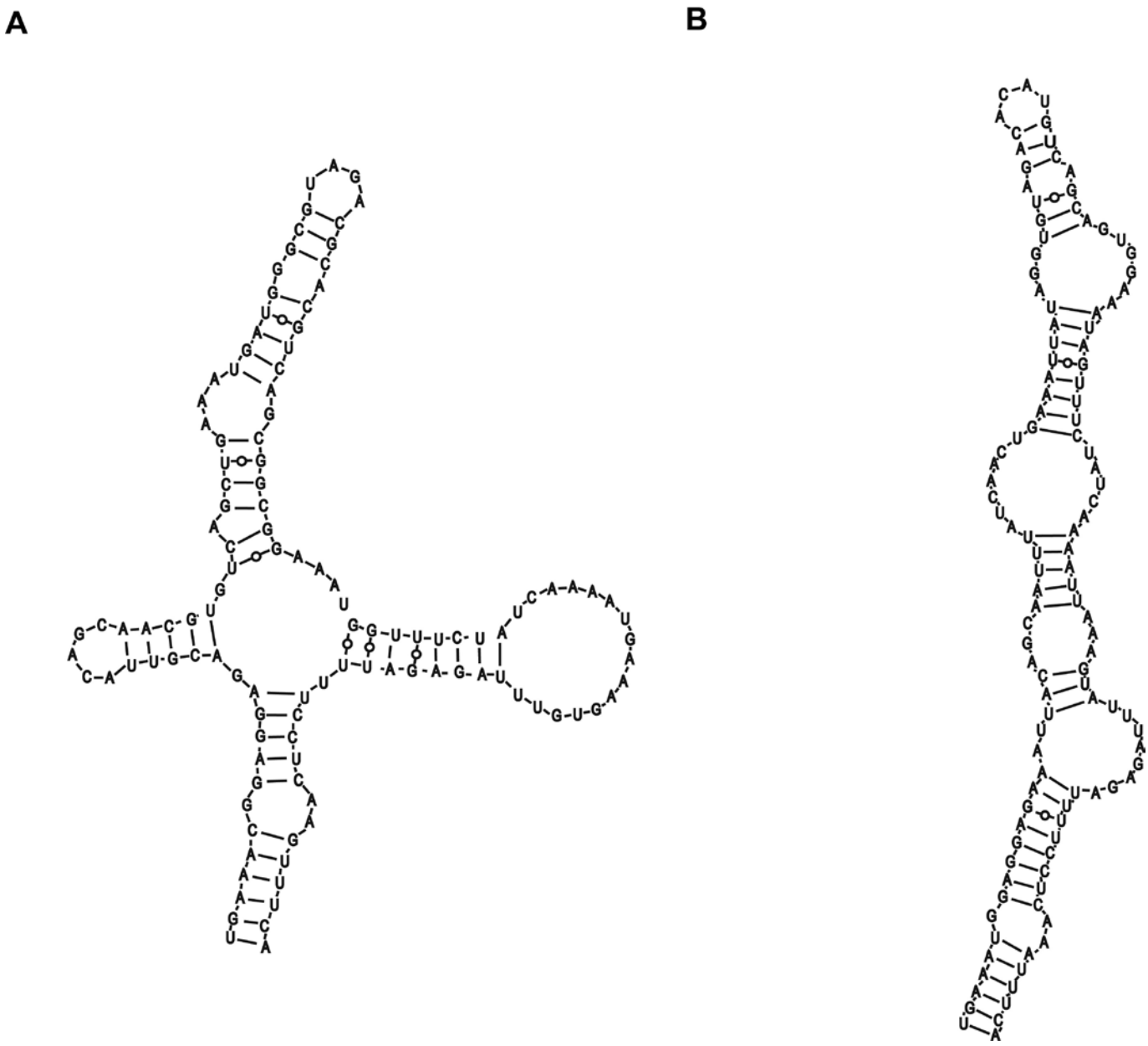


Figure 1. The secondary structures for the human HAR1 RNA (A) and the chimpanzee HAR1 (B). Notice that the difference in sequence results in a molecule with a significantly different shape. Based on their assumption of universal common ancestry, evolutionists believe this region of the genome underwent rapid evolution after humans split from chimps. Figure was adapted from Beniaminov, et al., 2008.

biochemical investigation performed several years after the original discovery of the gene (Beniaminov et al., 2008). The chimpanzee HAR1 RNA adopts a long hairpin structure, while the human HAR1 RNA forms a completely different cloverleaf structure. These dramatically different configurations are clearly associated with taxonomic specificity and function.

Other HAR Discoveries

At about the same time the discovery of HAR1 was being announced, several other reports were published describ-

ing larger genome-wide investigations of accelerated noncoding regions in humans and other vertebrates (Pollard et al., 2006a; Prabhakar et al., 2006). Prabhakar et al. (2006) compared the conserved noncoding regions (CNS) in humans, chimpanzees, and mice with the result that none of the overall patterns across lineages conformed to the grand evolutionary paradigm (inferred evolutionary trees). They also found that the CNS regions were heavily enriched near neuronal cell adhesion genes (cadherins, protocadherins, contactins, and neurologins) in chimpanzees and humans, but they were not in mice—a clear anomaly for the overall mammalian evolutionary model. Furthermore, disparate evolutionary results were obtained between humans and chimpanzees, to which the authors responded, "This suggests independent accelerated evolution of neuronal cell adhesion functions in the human and chimpanzee lineages" and "It is unlikely that acceleration of neuronal adhesion CNSs in humans and chimpanzees resulted in the same neuronal phenotypes, because the CNSs accelerated in the two lineages are largely disjoint and would therefore have had different consequences for brain development and cognitive function." In the end, they finally identified 992 CNS regions that were human-specific and attributed to advanced neural capabilities in humans versus other primates.

In the study by Pollard et al. (2006a), the researchers focused specifically on 202 carefully selected candidate regions they claimed had been under strong negative selection, which is defined as the removal of alleles that are deleterious (also referred to as purifying selection). However, it should be noted that selection is not actually observed in cases like this but merely inferred based on the variability of the compared sequences in question. These regions are essentially nonvariant in humans but are significantly different from their counterparts in chimpanzees. Thus, it is thought they evolved quickly and then became indispensable to the human lineage and further evolution was shut down (constrained) in these regions due to the newly acquired functional importance of the sequence. This is essentially the mindset of the evolutionist in evaluating such sequences in a comparative sense.

The closest genes to these CNS regions in the Pollard et al. (2006a) study were enriched for transcription factors, DNA-binding proteins, and regulators of nucleic acid metabolism; they were shown to be statistically correlated with high levels of association to cellular processes involved with development, neurogenesis, and morphogenesis.

Functionality of HARs

To help determine the functionality of HAR sequences, a recent study compiled a comprehensive list of 2,649 noncoding HARs, combining data from over five studies (Capra, et al., 2013). They then determined functionality for these regions using data from the ENCODE project for transcription factor binding, histone modifications, and other indicators of chromatin state. They also analyzed positional data to determine the genomic landscape in which these sequences were situated. Using this combinatorial data, they found that at least 30% were clearly functional enhancer elements, with more than half (~60%) of the elements showing enhancer activity in at least one type of cellular context. Thus, well over half of these types of sequences appear to function as enhancers.

Enhancers are short 50 to 1500 bp regions that bind with transcription factors to activate transcription of a gene (Capra, et al., 2013). They are generally cis-acting, and can be located up to 1 million bp away from a gene that they regulate, upstream or downstream from the gene's start site and in either the plus or minus strand orientation. Over 40,000 enhancers have been catalogued in the human genome, and many are related to developmental processes (Andersson et al., 2014). Enhancer HARs have been found to be enriched in both intergenic regions across the genome and intragenic regions inside introns (Capra et al., 2013). The HARs in the Capra et al. (2013) study were on average 257 bp long, and most were within 1 Mb of a known gene, with 19% of these genes encoding transcription factor binding sites. So clearly these are important regulatory sequences in the overall scheme of gene and genome regulation.

Interestingly, the researchers of the Capra et al. (2013) study also tested a small number of enhancers from both human and chimpanzees in transgenic mice. While this effort was not exhaustive in scope, a significant number of enhancers from both human and chimpanzee drove markedly different expression patterns in developing mouse embryos, indicating significant differences in functionality. At present, ten different HAR sequences have been tested in functional assays such as this in a variety of studies. Most were implicated in brain development, while two enhancer HARs were implicated in limb and eye development (Kamm et al., 2013a; Kamm et al., 2013b; Lindblad-Toh et al., 2011; Pollard and Franchini, 2015; Rossant, 2015; Sumiyama and Saitou, 2011). Of course, a limitation for studies like this (testing foreign constructs in transgenic mice) is that they cannot truly recapitulate the true function of a human or chimpanzee DNA regulatory element - they can only show how differences in the sequence produce different functional outcomes, and in some cases, what types of tissues their expression may correspond with (Pollard and Franchini, 2015).

The developmental process itself is orchestrated through complex regulatory networks that are tightly regulated and highly constrained (Davidson and Erwin, 2006). All types of DNA sequences, both developmental genes (e.g., transcription factors) and regula-

tory sequences (like HARs) play major roles in development. Transcription factor genes are highly pleiotropic. In other words, they participate in multiple independent processes, both spatially and temporally. In contrast, noncoding regulatory sequences, such as enhancers, tend to function in a more limited number of cell types and processes. They also tend to operate more in an additive manner—combining together to control the complex expression patterns of developmental genes such as transcription factors (Noonan and McCallion, 2010).

Evolutionists seem to think this highly efficient, yet complex system of regulatory and developmental gene modules is somehow conducive to evolution (Carroll, 2008), despite the fact that the evolutionary model cannot account for their origin and disrupting these sequences often leads to serious problems, including catastrophic system failure. The most obvious and parsimonious explanation is that this type of complex modularity in code is analogous to human-engineered computer software that is both modular and often object oriented in its construction, where methods (functions) can be called in an additive fashion to instances of an object, thereby controlling and altering its output in the overall program. The ingenious design patterns in the genome are truly spectacular, but the significance of the implications are generally missed by those with the mindset of an evolutionist entrenched in naturalistic thinking.

Deleted Accelerated Regions in Humans?

Not only are the presence of HARs an enigma for the evolutionary paradigm, but so is the absence of such regions when comparing taxa. One must keep in mind that within the evolutionary mindset, these regions are allegedly under strong selective constraint and thus differ very little in their sequence between taxa. Thus, their sudden "disappearance" from a genome in the grand evolutionary tree of life is difficult to account for.

In a large genome-wide survey for highly conserved sequences absent in human but present in chimpanzee and other mammals, researchers found 510 such sequences, all of which (except for one) mapped to noncoding regions of the human genome (McLean et al., 2011). Several of these allegedly deleted regions in humans corresponded to apparent enhancer elements present in the genomes of other mammals. The conserved chimp and mouse elements, along with deletions of them, were tested in transgenic mice. It was found that in transgenic mouse embryos, one of the deletions removed sensory vibrissae (tactile hair on the head, e.g., whiskers) and a penile spine enhancer element from a homolog to the human androgen receptor gene. The alleged deletion of this element in humans is quite large and corresponds to about 60,000 base pairs. Another supposed deletion was found to correspond to the removal of a forebrain subventricular zone (paired brain structure situated throughout the lateral walls of the lateral ventricles) enhancer element in transgenic mice.

This original study of these two specific highly conserved enhancer elements (present in other mammals but mysteriously missing in humans) were followed up several years later in another study (Reno et al., 2013). Using a combination of large-scale database sequence analyses and direct DNA analysis of the genomes in question, researchers demonstrated that the penile spine/vibrissa enhancer element was missing in all human genomes surveyed, and also in the archaic human genomes of Neandertal and Denisovan, but present in DNA samples of chimpanzees and the other great apes and other primates that exhibit some form of penile spine and facial vibrissae. The other 508 conserved elements supposedly deleted during evolution in a common ancestor of humans and chimps remain to be functionally characterized.

Another major evolutionary anomaly with overall patterns of these conserved noncoding elements in regard to their alleged mysterious deletion in major animal lineages is that the patterns are erratic and the supposed sudden absence of these elements are said to represent "independent losses" and are "not uniform" (Hiller et al., 2012). In other words, they do not form consistent evolutionary trees regarding their presence and absence across lineages. Hiller et al. (2012) explained the majority of these aberrant patterns by claiming that many of the lost elements were slightly less evolutionarily constrained and shorter and thus must have been less pleiotropic. Enhancer elements for the most part do appear to be less pleiotropic on average than protein-coding developmental genes (Carroll, 2008; Wray, 2007). But this is not really a satisfactory reason for their erratic presence or absence across major lineages, given their functional importance and the alleged evolutionary constraint ascribed to them.

Materials and Methods

To supplement the literature review in this report and to fill in a glaring gap within the HAR research community, the phylogenetic analysis of 105 different HAR sequences was undertaken for the following taxa: human, chimpanzee, gorilla, orangutan, macaque, mouse, elephant, cow, and chicken. The approach to acquiring the data was as follows: (1) Using url links at <docpollard.com/HARs.html>, each individual HAR sequence was followed to its respective hg17 "Vertebrate Multiz Alignment & Conservation" view at the UCSC genome browser (http://genome.ucsc.edu). (2) I then went to "View" then "Other genomes (Convert)"and used the more current hg19 version for my alignment data (adjusting the browser view for the species listed above). (3) I clicked on the

alignment link for each respective HAR gene and downloaded the subsequent alignment view as a plain text file. (4) I processed each downloaded text file with a Python script I wrote that converted it into standard FASTA file format.

The phylogenetic analysis pipeline was performed as follows: (1) The MUSCLE v3.8.31 program (Edgar, 2004) was used on the set of 105 FASTA files produced as described above, yielding multiple DNA alignment output files in FASTA format (MUSCLE default parameters). (2) The MUSCLE program was used again to produce neighbor-joining trees (parameters: -maketree, -cluster neighbor joining). (3) These individual tree files were further text-processed and combined into a single multitree specialized Newick-style file required by the tree comparison program topd_v3.3.pl (Puigbo et al., 2007). Steps 1 through 3 were performed via a Python pipeline script written by this author with MUSCLE being employed as system calls within Python. The resulting Newick-style, multitree file processed with the same Python pipeline script was analyzed with topd_v3.3.pl but also evaluated for commonalities in topology by basic UNIX shell programs such as uniq and grep, the latter was employed with a variety of different regular expressions for pattern matching. Phylogenetic trees, including those for this publication were drawn and printed to file using the Phylodendron Phylogenetic tree printer program (http://iubio.bio.indiana.edu/treeapp/treeprint-form.html). The two python scripts for processing the UCSC "Vertebrate Multiz Alignment & Conservation" text files and implementing the MUSCLE-tree pipeline, along with the FASTA files used in this study have been posted at github (https://github.com/jt-icr/har_code.git).

Results

The results of the phylogeny analyses of the 105 HAR sequences tested were inconsistent with the grand evolutionary paradigm, in complete accordance with all of the other data discussed above. Based on analyses with the topd_v3.3.pl program, which exhaustively compares tree topologies to each other using a variety of algorithms, there was no evidence for a unified evolutionary tree in this data set. These trees did not support the inferred evolutionary phylogeny for the species tested.

A sampling of the discordant trees is shown (incorporating genetic distance) in Figure 2. Most notable among these trees are those for HAR1 and HAR2, several of the best-studied HAR genes that are also noted for their evolutionarily unsupportive sequences (alleged acceleration in humans compared to chimpanzees). For HAR1, human and mouse cluster together in the same branch, as does elephant and chicken. The tree for HAR2 likewise is completely discordant with evolution, as human clusters with elephant.

As a whole, the different HAR genes gave widely different topologies. This is frequently observed and is attributed to incomplete lineage sorting, a rescuing device used by evolutionists to explain incongruent data. This type of evolution-negating pattern has been a common finding of studies analyzing many different genes, genomics regions, or even protein sequences (Degnan and Rosenberg, 2009; Hobolth et al., 2011; Pisani et al., 2012; Tomkins and Bergman, 2013). It appears the phylogenetic discordance for HAR sequences greatly exceeds that for other types of regions, such as protein-coding gene exons.

Even when analyzing subtrees within the data set, humans and chimpanzees clustered together on the same branch only on 15 occasions (14% of the trees). Gorilla and human clustered together in only 8 instances, and orangutans, supposedly more distant to humans than gorillas, clustered directly with human on 11 trees. Furthermore, a two-branch cluster with human and at least two apes (e.g. "[chimp, human] gorilla," "[gorilla, human] chimp," etc.) was only seen on 13 occasions. In other words, if two great apes occupied a branch with humans, it was typical for the other to be located on a completely separate branch.

Summary

Human accelerated regions (HARs) are noncoding DNA sequences in the genome that, according to evolutionary reasoning, changed very little over the course of animal evolution but mysteriously and quite suddenly experienced a "burst" of change since the alleged divergence of humans from chimpanzee. These HAR-type sequences also appear suddenly in assumed vertebrate lineages with no prior evolutionary history, while others disappear and then reappear.

In humans and several other mammals, many of these HARs are being functionally characterized as enhancer elements, developmental gene regulatory elements, and even noncoding RNA genes. Many of them are also associated with a wide variety of important neurological traits unique to the humans. Evolutionists claim that the lack of variation in these sequences among other animals is due to "conserved function." Of course, very little is actually known about what these sequences are doing in the different kinds of animals in which they are found. In reality, we are just beginning to discover what they are doing even in humans.

In addition to the alleged "accelerated" evolution of these sequences within the human genome, this study shows that HAR genes show a pattern inconsistent with evolutionary predictions about the common ancestry of vertebrate lineages; this pattern is typically explained away as incomplete lineage sorting. The experimental data presented in this report shows that these alleged highly conserved sequences are discordant with classic evolutionary phylogenetic analyses. The analysis of 105

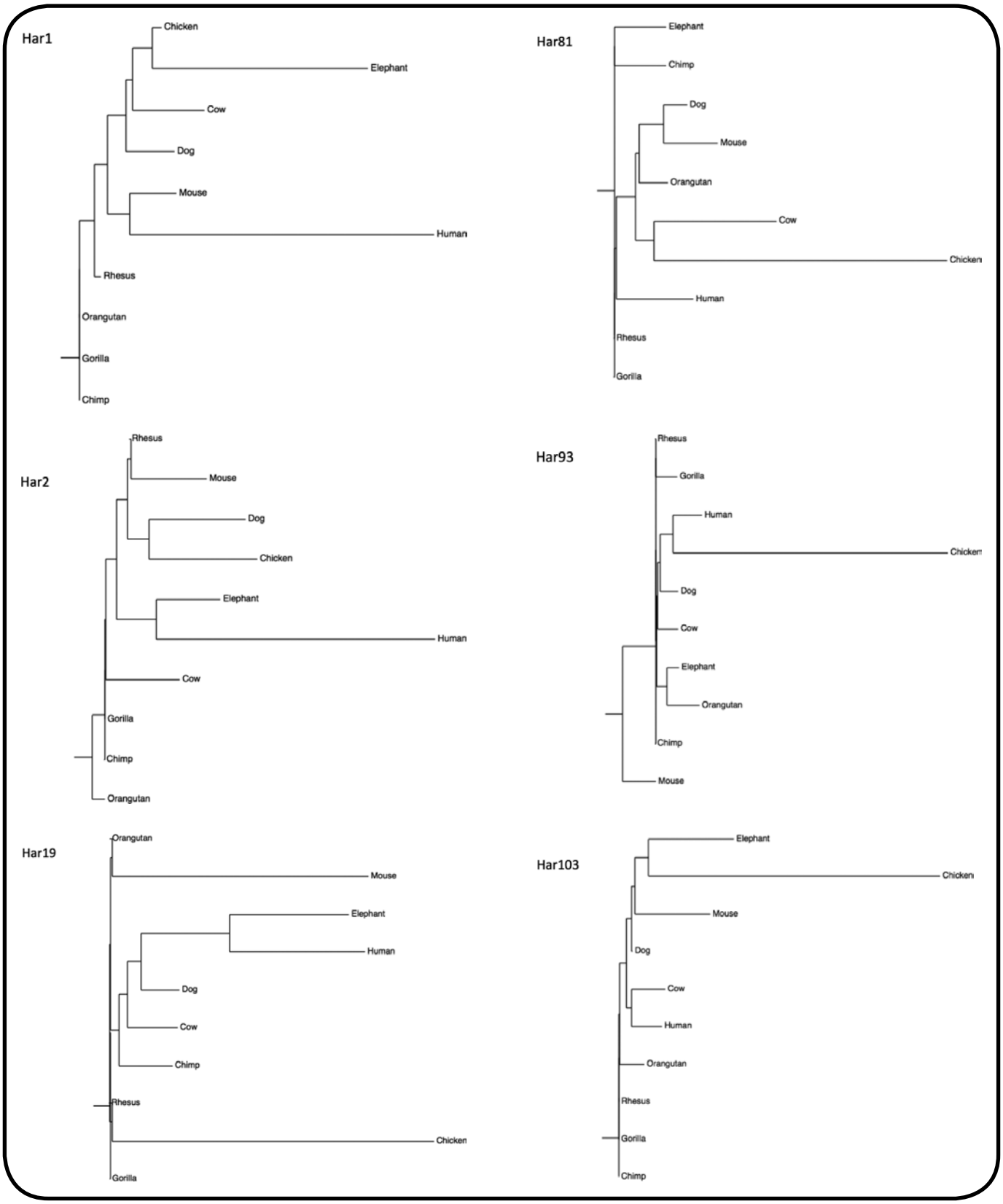


Figure 2. A selection of six different neighbor–joining trees from the 105 HAR phylogenies produced using MUSCLE program. Philodendron was used to draw the trees. Notice how inconsistent the results are, showing that the hypothesis of universal common ancestry is not supported by these genetic data.

different HAR genes from 10 different vertebrate taxa, including humans and the great apes, show extremely discordant evolutionary trees.

So what can we make of all this evolutionarily incongruent data surrounding HARs? Cleary, the most parsimonious answer is that they represent designed functional mammalian genetic elements that encode the novel phenotype of mankind. This is consistent with humans being uniquely created in the image of God, as clearly stated in the Bible. There is no evidence that these human-specific sequences evolved at any level or that they experienced a "burst of changes in humans since divergence from chimpanzees."

The standard explanation for HARs is also clearly falsified by recent research that has shown that for any mammalian species there is a *profound waiting time problem* associated with establishing new traits that require multiple new mutations (Sanford et al., 2015). Even establishing two codependent mutations in a hominin population is extremely problematic—requiring tens of millions of years. Since HAR genes are different at many nucleotide positions, the hypothesis that HAR genes arose very suddenly in just a few million years due to accelerated evolution is not even remotely credible.

In Psalm 139:14, 16, it is stated: "I will praise thee; for I am fearfully and wonderfully made ... and in thy *book* all my members were *written*." The Hebrew word for book is *siphrah*, which means a writing or document and by implication, a book, letter, or scroll. The Hebrew word for written is kâtab, which means to write, describe, inscribe, prescribe, or subscribe. We now know from the study of genetics and genomics that the genome is, in fact, a highly complex multidimensional document written in multiple codes and languages that we are now only beginning to understand (Tomkins, 2015a). Needless to say, this kind of handiwork far exceeds the abilities of even humans to engineer (or even fully understand). It clearly points to the Creator described in the Bible.

References

Andersson, R., C. Gebhard, I. Miguel-Escalada, et al. 2014. An atlas of active enhancers across human cell types and tissues. *Nature* 507(7493): 455–461.

Beniaminov, A., E. Westhof, and A. Krol. 2008. Distinctive structures between chimpanzee and human in a brain noncoding RNA. *RNA* 14(7): 1270–1275.

Bergman, J., and J. Tomkins. 2012. Is the human genome nearly identical to chimpanzee? A reassessment of the literature. *Journal of Creation* 26:54–60.

Capra, J.A., G.D. Erwin, G. McKinsey, J.L.R. Rubenstein, and K. Pollard. 2013. Many human accelerated regions are developmental enhancers. *Philosophical Transactions Royal Society London Biological Sciences* 368(1632): 20130025.

Carroll, S.B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134(1): 25–36.

Chimpanzee Genome Sequencing Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437(7055): 69–87.

Davidson, E.H., and D.H. Erwin. 2006. Gene regulatory networks and the evolution of animal body plans. *Science* 311(5762): 796–800.

De, S., S.A. Teichmann, and M.M. Babu. 2009. The impact of genomic neighborhood on the evolution of human and chimpanzee transcriptome. *Genome Research* 19(5): 785–794.

Degnan, J.H., and N.A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* 24(6): 332–340.

Dickel, D.E., A. Visel, and L.A. Pennacchio. 2013. Functional anatomy of distant-acting mammalian enhancers. *Philosophical Transactions Royal Society London Biological Sciences* 368(1620): 20120359.

Ebersberger, I., P. Galgoczy, S. Taudien, S. Taenzer, M. Platzer, and A. von Haeseler. 2007. Mapping human genetic ancestry. *Molecular Biology Evolution* 24(10): 2266–2276.

Edgar, R.C. 2004. Muscle: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.

Formenti, D., S. Garagna, G.F. De Stefano, and C. Pellicciari. 1983. Variazioni del contenuto nucleare in DNA negli hominoidea. *Antropologia Contemporanea* 6:219–224.

Hiller, M., B.T. Schaar, and G. Bejerano. 2012. Hundreds of conserved non-coding genomic regions are independently lost in mammals. *Nucleic Acids Research* 40(22): 11463–11476.

Hobolth, A., O.F. Christensen, T. Mailund, and M. Schierup. 2007. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden markov model. *PLoS Genetics* 3(2): e7.

Hobolth, A., J.Y. Dutheil, J. Hawks, M. Schierup, and T. Mailund. 2011. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Research* 21(3): 349–356.

Kamm, G.B., R. Lopez-Leal, J.R. Lorenzo, and L.F. Franchini. 2013a. A fast-evolving human npas3 enhancer gained reporter expression in the developing forebrain of transgenic mice. *Philosophical Transactions Royal Society London Biological Sciences* 368(1632): 20130019.

Kamm, G.B., F. Pisciottano, R. Kliger, and L.F. Franchini. 2013b. The developmental brain gene npas3 contains the largest number of accelerated regulatory sequences in the human genome. *Molecular Biology and Evolution* 30(5): 1088–1102.

Koop, B.F., M. Goodman, P. Xu, K. Chan, and J.L. Slightom. 1986. Primate globin DNA sequences and man's place

among the great apes. *Nature* 319(6050): 234–238.

Lindblad-Toh, K., M. Garber, O. Zuk, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478(7370): 476–482.

McLean, C.Y., P.L. Reno, A.A. Pollen, et al. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* 471(7337): 216–219.

Noonan, J.P., and A.S. McCallion. 2010. Genomics of long-range regulatory elements. *Annual Review of Genomics and Human Genetics* 11:1–23.

Patterson, N., D.J. Richter, S. Gnerre, E. Lander, and D. Reich. 2006. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 441(7097): 1103–1108.

Pellicciari, C., D. Formenti, C.A. Redi, and M.G. Manfredi Romanini. 1982. DNA content variability in primates. *Journal of Human Evolution* 11:131–141.

Pellicciari, C., D. Formenti, M. Zuccotti, R. Stanyon, and M.G. Manfredi Romanini. 1988. Genome size and constitutive heterochromatin in hylobates muelleri and symphalangus syndactylus and in their viable hybrid. *Cytogenetics and Cell Genetics* 47(1–2): 1–4.

Pellicciari, C., E. Ronchetti, D. Formenti, R. Stanyon, and M.G. Manfredi Romanini. 1990a. Genome size and c-heterochromatic DNA in man and the African great apes. *Human Evolution* 5(3): 261–267.

Pellicciari, C., E. Ronchetti, R. Tori, D. Formenti, and M.G. Manfredi Romanini. 1990b. Cytochemical evaluation of c-heterochromatic-DNA in metaphase chromosomes. *Basic and Applied Histochemistry* 34(1): 79–85.

Pisani, D., R. Feuda, K.J. Peterson, and A. Smith. 2012. Resolving phylogenetic signal from noise when divergence is rapid: A new look at the old problem of echinoderm class relationships. Molecular Phylogenetics and Evolution 62(1): 27–34.

Pollard, K.S., S.R. Salama, B. King, et al. 2006a. Forces shaping the fastest evolving regions in the human genome. *PLoS Genetics* 2(10): e168.

Pollard, K.S., S.R. Salama, N. Lambert, et al. 2006b. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443(7108): 167–172.

Pollard, K.S., and L.F. Franchini. 2015. Can a few non-coding mutations make a human brain? *Bioessays* 37: 1054–1061.

Prabhakar, S., J.P. Noonan, S. Paabo, and E. Rubin. 2006. Accelerated evolution of conserved noncoding sequences in humans. *Science* 314(5800): 786.

Prado-Martinez, J., P.H. Sudmant, J.M. Kidd, et al. 2013. Great ape genetic diversity and population history. *Nature* 499(7459): 471–475.

Puigbo, P., S. Garcia-Vallve, and J.O. McInerney. 2007. Topd/fmts: a new software to compare phylogenetic trees. *Bioinformatics* 23(12): 1556–1558.

Reno P.L., C.Y. McLean, J.E. Hines. T.D. Capellini, G. Bejerano, and D.M. Kingsley. 2013. A penile spine/vibrissa enhancer sequence is missing in modern and extinct humans but is retained in multiple primates with penile spines and sensory vibrissae. *PLoS One* 8(12):e84258.

Rossant, J. 2015. Mouse and human blastocyst-derived stem cells: vive les differences. *Development* 142(1): 9–12.

Sakabe, N.J., D. Savic, and M.A. Nobrega. 2012. Transcriptional enhancers in development and disease. *Genome Biology* 13(1): 238.

Sanford, J., W. Brewer, F. Smith, and J. Baumgardner. 2015. The waiting time problem in a model hominin population. *Theoretical Biology and Medical Modeling* 12:18.

Seuanez, H.N., A.D. Carothers, D.E. Martin, and R.V. Short. 1977. Morphological abnormalities in spermatozoa of man and great apes. *Nature* 270(5635): 345–347.

Sumiyama, K., and N. Saitou. 2011. Loss-of-function mutation in a repressor module of human-specifically activated enhancer hacns1. *Molecular Biology and Evolution* 28(11): 3005–3007.

Tomkins, J. 2011. How genomes are sequenced and why it matters: implications for studies in comparative genomics of humans and chimpanzees. *Answers Research Journal* 4:81–88.

Tomkins, J., and J. Bergman. 2012. Genomic monkey business—estimates of nearly identical human-chimp DNA similarity re-evaluated using omitted data. *Journal of Creation* 26:94–100.

Tomkins, J., and J. Bergman. 2013. Incomplete lineage sorting and other 'rogue' data fell the tree of life. *Journal of Creation* 27(3): 84–92.

Tomkins, J. 2015a. Extreme information: biocomplexity of interlocking genome languages. *Creation Research Society Quarterly* 51:187–201.

Tomkins, J.P. 2015b. Documented anomaly in recent versions of the blastn algorithm and a complete reanalysis of chimpanzee and human genome wide DNA similarity using nucmer and lastz. *Answers Research Journal* 8:379–390.

Wray, G.A. 2007. The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics* 8(3): 206–216.

Creation Research Society Quarterly 2016. 52:265–274.

Adaptive Genetic Changes by Design: A Look at the DNA Editing by Activation-induced Cytidine Deaminase (AID)

Jean Lightner*

Abstract

According to evolutionary thinking, adaptive genetic changes are the result of random (non-purposeful) mutations and natural selection. While creationists do not need to account for the assumed changes that turn microbes into people, our model certainly points to a considerable amount of adaptive change that has occurred within created kinds. The naturalistic mechanisms proposed by evolutionists appear woefully inadequate to account for these.

A look at the immune system reveals several different enzymes that are used to edit DNA; one of them is activation-induced cytidine deaminase (AID). AID is involved in gene conversion, somatic hypermutation, and class-switch recombination in B lymphocytes. While each of these begins with AID converting a cytosine residue to a uracil residue, the different outcomes are a function of different proteins being recruited to process the lesion. Since the activity of AID could be disastrous if not kept in its proper place, it is well regulated and tightly controlled at many levels.

The well-designed DNA editing function of AID and other proteins in the immune system give reason to believe that adaptive alleles in various populations have similarly arisen by the providence of God, the Great Designer, and not by the naturalistic mechanisms proposed by evolutionists.

Introduction

According to popular evolutionary thinking, often referred to as neo-Darwinism or the modern evolutionary synthesis, the source of variation upon which natural selection supposedly acts arises from mutation. While the term "mutation" has changed over the last hundred years or so, it now generally refers to a change in the DNA sequence, primarily caused by unrepaired errors during replication (Mayr, 2001, pp. 96–98, 279–280). It is insisted that there is no teleology involved; that is, there is no design or purpose underlying mutations (Mayr, 2001, pp. 119–120, 275). They are believed to arise by chance and be random with respect to the needs of the organism (Huxley, 2010, p. 54). Natural selection is the mechanism given credit for rare adaptive genetic changes becoming fixed in various populations of organisms

* Jean Lightner, D.V.M., jklightner@gmail.com
Accepted for publication April 26, 2016

(Rensch, 1980, pp. 296–298; Mayr, 2001, 119–120).

Creationists approach the scientific evidence from a different vantage point. We reject the assumption of universal common ancestry, and instead recognize that creatures were created as various kinds (Genesis 1:11–12, 21, 24–25). Humans were created separately from all other animals (Genesis 1:26–28; 2:7, 19–24). In the process of analyzing biological data within a biblical framework, many have questioned two additional assumptions of neo-Darwinism: that mutations are always random errors without purpose, and that natural selection can really explain the almost magical transformations and adaptations in populations that evolutionists claim (Purdom and Anderson, 2009; Terborg, 2008; Lightner, 2015).

There is strong evidence that adaptive genetic changes do occur. The biblical history describing the global Flood makes it clear that limited genetic variability was present immediately after that event. This is especially true of unclean land animals, where only two individuals survived the Flood on the ark, but even clean animals and birds on the ark had a drastically reduced population. Certainly, much more genetic and phenotypic variability is present today compared to the time of the Flood (Lightner, 2006, 2009a; Wayne and vonHoldt, 2012). In some groups, such as birds, there is astounding diversity that has arisen within created kinds (Lightner, 2010b). Particularly impressive are the adaptive radiations of birds inhabiting islands, such as the radiations of the vangas of Madagascar, the honeycreepers in Hawaii, and the finches in the Galápagos (Reddy et al., 2012; Jønsson et al., 2012; Lerner et al., 2011; Lamichhaney et al., 2015).

A recent creationist review of high altitude adaptation shows that adaptation is a complex, multilevel process that ranges from short-term physiological adjustments in the individual to new adaptive alleles in populations (Lightner, 2014). Creationists have proposed various mechanisms for genomic change beyond the well-known shuffling that occurs during homologous recombination (i.e., crossing over and gene conversion in meiosis). Considerable attention has been given to transposable elements (Wood, 2002; Terborg 2009a, 2009b; Shan, 2009). Many transposable elements contain the genetic instructions for their own movement, and certain conditions seem to activate their movement (e.g., stress, hybridization). They can change the sequence of a gene, alter the regulation of one or more genes, and/or be involved in chromosomal rearrangements (Belyayev, 2014).

While transposable elements certainly appear to play a role, many detailed studies of genetic differences underlying phenotypic diversity have yielded few examples of where transposable elements appear to be involved in the genetic mutations that were identified (Lightner 2008, 2009b, 2010a). Therefore, within the creation model, a reasonable prediction is that other mechanisms are involved in many of these DNA changes. It has been pointed out that DNA changes in B cells are a normal part of mounting an antibody response and that all the necessary biological information required to induce appropriate variation is coded in the genome (Terborg, 2009a). Therefore, DNA editing enzymes involved in the immune system are worth considering in more detail, and recent reviews highlight advances in our understanding of the underlying mechanisms involved in the essential functions they perform (Zan and Casali, 2013; Kumar et al., 2014; Matthews et al., 2014; Moris et al., 2014; Chandra et al., 2015).

DNA Editing and the Immune System

Two enzymes known to be involved in DNA editing within the immune system are APOBEC3 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3) and activation induced cytidine deaminase (AID). They belong to the AID/APOBEC family of DNA and RNA editing enzymes, which have important roles in a variety of important functions (Moris et al., 2014). This family is unique in possessing the ability to deaminate cytidine (C; the cytosine residue in RNA) or deoxycytidine (dC; in DNA) to uridine (U) or deoxyuridine (dU) (Figure 1). In many species, APOBEC3 varies in copy number and is polymorphic; it restricts the replication of many exogenous viruses and endogenous transposable elements (Harris and Dudley, 2015). At least some of this activity is based on its DNA editing ability, which essentially mutates the viral genome, destroying its ability to replicate.

While the best-known roles of APOBEC3 involve innate immunity, AID plays critical roles in various steps of adaptive immunity. Adaptive immunity involves an ingenious design that enables creatures to adapt to the challenges in their environment. Organisms are constantly exposed to a myriad of

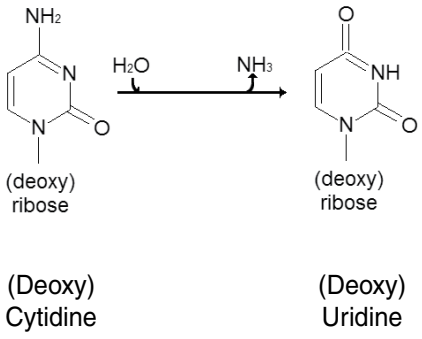


Figure 1. Activation-induced cytidine deaminase (AID) deaminates the cytosine residue deoxycytidine (dC) to deoxyuridine (dU) in DNA to enable adaptive immune responses. Other members of the AID/APOBEC family also catalyze this reaction in DNA or RNA for other purposes.

potentially harmful microorganisms, parasites, and toxins. They need to be able to identify them and properly dispose of them as necessary. Rather than being front-loaded with the exact code for every antibody to every possible antigen that could be encountered, the adaptive immune system is strategically designed to manufacture highly specific antibodies that can be used in several different contexts to effectively deal with potential pathogens.

The portion of the genome used for antibody (immunoglobulin, Ig) formation already contains some variability in many species. For example, in humans and mice, there are a number of different variable (V), joining (J), and diversity (D) regions coded on the DNA. Through V(D)J recombination, the recombination activating gene enzymes (RAG1 and RAG2) initiate double-stranded breaks in the DNA that are repaired to bring a single V, D, and J segment in apposition with each other (Jung et al., 2006). Other steps in Ig formation that involve DNA sequence modification are gene conversion, somatic hypermutation (SHM), and class-switch recombination (CSR). Each of these three steps uses AID to initiate the genomic changes (Arakawa et al., 2004; Matthews et al., 2014).

Gene conversion in lymphocytes was first described in the chicken, which has only one V region for both the light and heavy chain loci involved in Ig formation. However, it was found that there were numerous pseudogene V regions upstream that provide templates for intrachromosomal gene conversion, copying nucleotide tracts from the pseudogenes onto the V region to increase Ig diversity. In the rabbit, lymphocytes also undergo gene conversion to increase variability; however, some of the upstream sequences are potentially functional. Interestingly, in both species, this process can be used to diversify the primary antibody repertoire, or further increase diversity in an antigen-specific immune response (Lanning and Knight, 2015).

Somatic hypermutation (SHM) involves the rapid introduction of mutations, primarily single nucleotide changes, into the complementarity determining regions (CDRs) of the recombined V region. The CDRs code for the portion of the Ig molecule that contacts the antigen. SHM is best known for its role in an antigen-specific immune response, and there is a designed mechanism providing for the selection of B cells expressing Ig with the greatest affinity to the antigen. SHM enables the body to rapidly produce highly effective Ig to any conceivable antigen that is encountered from the more limited diversity of the primary antibody repertoire (Matthews et al., 2014).

Once an effective antibody has been produced, there is a need to use it in several different contexts to effectively deal with an infection. That is the purpose of class switch recombination (CSR); it is said to change the effector functions. To switch the class of Ig produced, a DNA segment needs to be excised to place the V(D)J region before an exon coding for a different constant (C) region (Figure 2). There are switch (S) regions containing repetitive DNA that precede most of the C regions, and these S regions are targeted by AID to induce several double-stranded breaks so the intervening region can be removed (Matthews et al., 2014).

AID: Gene Structure and Regulation

In humans, the enzyme AID is encoded by the *AICDA* gene on chromosome 12. The gene spans 11 kb, consists of 5 exons, and is primarily expressed in B cells. However, in mice it has also been detected in oocytes, embryonic germ cells, and embryonic stem cells; additionally, it has been detected in normal human spermatocytes. Occasionally, AID has been associated with pathology, as it has sometimes been detected in cells affected by chronic inflammation or cancer (though generally not testicular cancer). Since off-site activity of this DNA editing molecule can be disastrous, its expression is tightly controlled on multiple levels (Barreto and Magor, 2011).

In addition to a promotor region known to bind four different transcription factors, several other regions are involved in controlling transcription of the gene. Intron 1 contains an enhancer/silencer region, where two transcription factors bind to repress the gene, and two different transcription factors bind to de-repress it. A third downstream region binds a transcription factor to maintain physiologic levels of AID expression. A fourth enhancer region is located upstream of the promotor. Recently, three more enhancers were identified further (up to 50 kb) from the *aicda* locus in mice (on chromosome 6; Chandra et al., 2015; Kumar et al., 2014).

Once the gene is transcribed, stability of the mRNA is affected by two different microRNA (miRNA) molecules. These miRNAs bind to the 3' UTR, the untranslated region following the portion of mRNA specifying the amino acid sequence. One miRNA is down-regulated during B cell activation, while the other is up-regulated. It is not surprising, therefore, that experimentally induced mutation of the 3' UTR of the AID mRNA resulted in spatiotemporal dysregulation of AID and off-site mutations (Chandra et al., 2015; Kumar et al., 2014).

Enzyme Structure

The enzyme is 198 amino acids long in humans and consists of a number of functional domains, some of which overlap (pleiotropy) and a few of which are not well characterized. Like other cytidine deaminases, the catalytic region of AID (amino acid positions 56–90) includes two cysteines and a histidine that

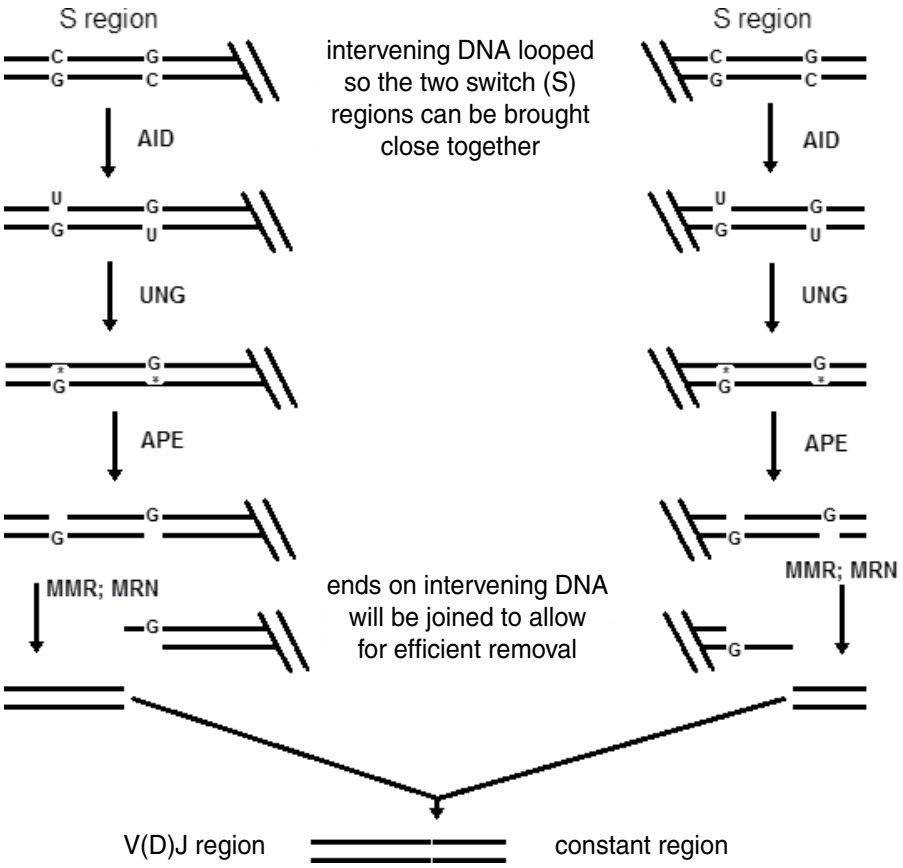


Figure 2. In class switch recombination (CSR), a segment of DNA must be removed to place the V(D)J region next to a different constant region. Activation-induced cytidine deaminase (AID) deaminates cytosine residues to uracil residues in the two switch (S) regions flanking the segment to be removed. The base excision repair enzyme uracil DNA glycosylase (UNG) removes the abnormal base leaving an abasic site. Apurinic/apyrimidinic endonuclease (APE) nicks the DNA at the abasic site. These single-stranded breaks can be converted to staggered double-stranded breaks. Mismatch repair (MMR) enzymes and the MRN complex (Mre11/Rad50/Nbs1) process the staggered ends so the V(D)J region can be joined to the constant region. The intervening DNA is circularized and removed.

coordinate a zinc ion to form the active site. Further downstream (113–123), though adjacent in the folded protein, is a critical hotspot recognition loop. This loop specifically targets a weak (W = A/T) nucleotide followed by a purine (R = A/G) in the 2' and 1' positions relative to the dC to be directed into the active site (Barreto and Magor, 2011; Nabel et al., 2014).

Interestingly, AID's preference for a DNA substrate appears to be related to the nucleotide's rotational conformation, sometimes known as sugar pucker. While AID targets a WRC motif (Figure 3), APOBEC3G (A3G) favors CCC. Experimental grafting of the recognition loop from one to the other will change the sequence specificity. When the AID loop was grafted into an A3G background, the chimeric enzyme was still efficacious in restricting effective HIV infection despite the difference in targeted sequence. In contrast, when the recognition loop of AID was changed, it adversely affected both SHM and CSR. Both the CDRs targeted in SHM and the S regions targeted in CSR are enriched with the WRC motif. This is accomplished within the CDRs by a preferential use of codons for serine (Ser), for example, that result in WRC hotspots, while in neighboring regions codons are preferentially used that do not create hotspots (Nabel et al., 2014; Kohli et al., 2010).

Among the other domains in the AID enzyme are a region required specifically for CSR (positions 190–198), another required for SHM (13–23), and a nuclear export signal (NES, 183–198). The latter plays an important role in maintaining AID in the cytoplasm, thus preventing it from damaging DNA when it has not been specifically recruited to provide an appropriate function (Barreto and Magor, 2011; Zan and Casali, 2013).

Keeping AID Where It Belongs

The hydrophilic residues in the NES (C terminal) portion of AID are essential for its active exclusion from the nucleus. It is retained in the cytoplasm, where it is complexed with other molecules to stabilize it until it is actively imported to the nucleus to perform its required function. It appears that there is a ubiquitin-dependent and independent pathway by which nuclear AID can be targeted for degradation (Barreto and Magor, 2011; Zan and Casali, 2013; Chandra et al., 2015). Phosphorylation of AID Ser3 is one factor that contributes to its degradation. Additionally, AID abundance in the nucleus is positively correlated with its catalytic activity (Matthews et al., 2014; Le and Maizels, 2015).

Somatic mutations in B cells are initiated during the G1 phase of the cell cycle. This is the growth phase that occurs prior to DNA replication (S phase), which precedes mitosis. When AID is experimentally sustained in the nucleus during the S phase, cell viability

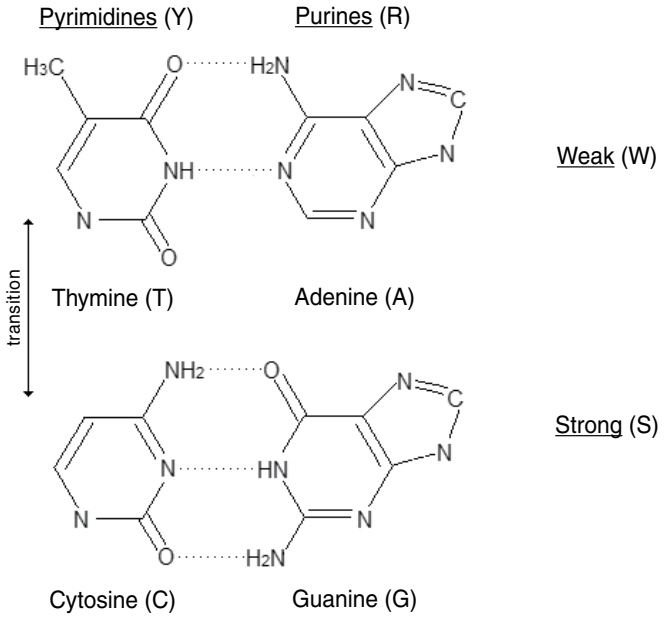


Figure 3. Nucleotides are classified as pyrimidines (Y) or purines (R) based on their ring structure. A mutation from one Y to another Y, or one R to another R is a transition; the ring structure stays the same. Mutations that result in a nucleotide with a different ring structure are called transversions. Nucleotides can also be classified as weak (W) or strong (S) based on the number of hydrogen bonds (2 or 3) in pairing. AID prefers the motif WRC (T or A; A or G; C)

is compromised; this explains why AID is normally rapidly degraded in the nucleus outside the G1 phase. It has been suggested that the nick left after the cellular machinery removes the foreign base, dU, may lead to double-stranded breaks if not repaired before replication. If so, this may account for the AID-dependent translocations characteristic of B-cell lymphomas where AID expression is no longer normally controlled (Le and Maizels, 2015).

Phosphorylation of AID Ser38 is necessary for somatic hypermutation (SHM) and class switch recombination (CSR), apparently to enable interaction with other molecules essential to these processes. In SHM, pS38-AID interacts with the single-stranded DNA (ssDNA) binding replication protein A (RPA), which stabilizes the ssDNA substrate that is the target of AID activity. In CSR pS38-AID recruits RPA and has been shown to interact with apurinic/apyrimidinic endonuclease (APE), which is required for making the breaks in the DNA. Interestingly, double-stranded breaks promote AID Ser38 phosphorylation, suggesting a positive feedback loop amplifies activity in S regions (Kumar et al., 2014; Matthews et al., 2014).

AID recruitment is transcription dependent. The transcripts through the V region (in SHM) or S region (CSR) are not translated, but in at least the latter case they are spliced. Deletion of a splice donor site was shown to interfere with CSR, suggesting that the transcripts might perform a regulatory function in some cases. During transcription, RNA polymerase II (Pol II) is stalled, and a factor involved in Pol II elongation and stalling, Spt5, has been shown to recruit AID. Additional adapter proteins have been found to recruit AID through their interaction with the abundant AGCT repeats (AGC being a subset of WRC) in the S region. Several other factors have been shown to be involved in recruiting AID as well. In fact, it has been commented that a surprisingly high number of cofactors are implicated despite the small size of the AID molecule, reflecting its tight regulation (Zan and Casali, 2013; Matthews et al., 2014; Chandra et al., 2015).

It is not difficult to understand how transcription allows AID access to the non-template strand; however, AID accesses both strands, which allows for the deaminated residues to be converted into double-stranded breaks for CSR. The RNA exosome complex has been shown to associate with AID and accumulate on S regions in an AID-dependent manner. This macromolecular complex removes and/or degrades nascent RNA on the template strand at stalled Pol II sites, exposing ssDNA for AID to access (Matthews et al., 2014; Chandra et al., 2015).

Epigenetic factors are also associated with AID recruitment. Methylated dCs make poor substrates for AID, and various histone modifications have been identified as playing a role in AID recruitment. A recent study evaluating patterns in both normal and off-site AID targeting found that regions enriched with chromatin modification typical of active enhancers, such as histone H3 acetylated at lysine 27 (H3K27Ac), as well as modifications typical of active transcription, such as trimethyl histone H3 lysine 36 (H3K36me3), mediate AID recruitment. Several transcription-factor binding sites are implicated in recruiting AID. Most AID targets are grouped within super-enhancers and regulatory clusters (Zan and Casali, 2013; Matthews et al., 2014; Chandra et al., 2015).

Diversity in Outcome via Different Repair Mechanisms

Despite the fact that gene conversion, SHM, and CSR in B cells all require

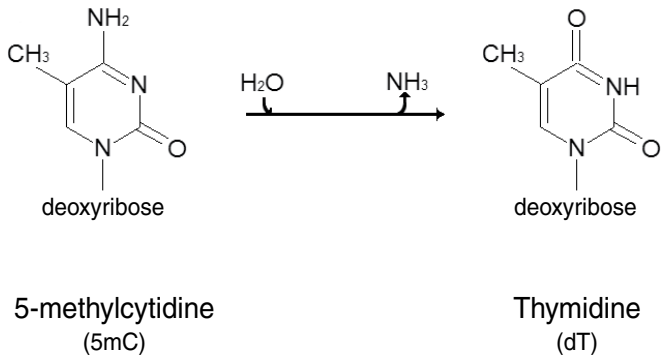


Figure 4. A methyl group (CH_3) can be added to deoxycytidine (dC; the cytosine residue in DNA) as an epigenetic tag to help regulate gene expression. At times the methyl group needs to be removed, and AID is sometimes involved. AID deaminates 5-methylcytidine (5mC) to thymidine (dT). This creates a T:G mismatch, which is subsequently repaired back to C:G, leaving an unmethylated cytosine residue

AID to initiate the process, the outcomes are very different. This is the result of recruiting a different array of proteins to process the dU lesion that AID creates.

In SHM, according to the current model, there are three possible pathways to repair the AID induced dU:dG mismatches. Replication prior to interaction with other repair enzymes results in transition mutations (CG → TA). Alternatively, removal of dU by the base excision repair (BER) enzyme uracil DNA glycosylase (UNG) prior to replication results in an abasic site. Subsequent repair during replication by error-prone DNA polymerases can lead to transition or transversion mutations. Otherwise, dU:dG mismatches can be processed by mismatch repair proteins, followed by filling in the gap with an error-prone polymerase, resulting in mutations at neighboring A:T residues and/or short indels (Matthews et al., 2014; Kumar et al., 2014).

It is important to recognize that error-prone polymerases are an essential part of the arsenal of polymerases used by cells to maintain genomic stability. They are specifically recruited to sites of DNA damage that the high processivity, high fidelity (i.e., fast, and accurate) polymerases cannot handle. In many (but not all) cases they accurately repair the lesions, though in the case of the immune system, they are recruited to induce changes (Saugar, et al., 2014; Yang, 2014).

According to the current model for CSR, dU introduced by AID is removed by the BER enzyme UNG (Figure 2). The abasic site is then converted to a ssDNA break by APE. A similar nick nearby on the opposite strand results in a staggered, double-stranded break. It has also been found that components of the mismatch repair pathway can act on dU:dG mismatches to form double-stranded breaks. These breaks are then repaired by nonhomologous end joining (Matthews et al., 2014; Kumar et al., 2014).

Theoretically, there are several ways the loose ends can be rejoined during CSR. For example, the intervening segment between the two S regions containing double-stranded breaks could be inverted, which would not result in a functional antibody. However, it appears there are certain features of the S region and AID designed to facilitate proper joining of the segments. The majority of the time the intervening segment is circularized, and the V(D)J region is correctly attached to the new C region (Dong et al., 2015).

Other Roles of AID

In addition to its roles in DNA sequence diversification in the immunoglobulin genes of B cells, AID appears to have other important functions. Methylation is a common epigenetic tag that helps define gene expression patterns essential to life. AID had been shown to deaminate 5-methylcytosine (5mC) to thymidine (dT), though the efficiency of this reaction is at least an order of magnitude lower than its normal substrate (Figure 4). When this reaction takes place, it results in a T:G mismatch that can be processed by glycosylases and downstream BER enzymes to restore an unmethylated C. Currently, there are conflicting conclusions on the relevance of this reaction in vivo based on studies. One recent summary suggests that there is no strong evidence for AID in genome-wide demethylation, but it appears to play a role in gene-specific demethylation that underlies cell differentiation (Ramiro and Berreto, 2015).

AID is also important in B-cell tolerance, and lack of the enzyme is associated with autoimmune disease. This is a rather paradoxical phenomenon, where humans lacking AID not only suffer from infections because they cannot mount a normal antibody response but also suffer autoimmune disease due to the inability to remove autoreactive B cells. In this role AID is expressed in immature B cells along with RAG2, though many details of how they eliminate autoreactive B clones remain to be elucidated (Cantaert et al., 2015).

Obviously, although AID has numerous crucial functions, loss of control over AID or the associated DNA repair

pathways can have disastrous results. Hypomethylation, point mutations, indels, and structural rearrangements are all features that are associated with cancer. Off-site activity of AID appears to be one factor that can contribute to carcinogenesis in certain malignancies such as lymphoma (Dominguez and Shaknovich, 2014; Pettersen et al., 2014).

Interestingly, in addition to mutations driving oncogenesis, cancer cells carry many thousands of passenger mutations not directly related to disease progression. With the increased availability of rapid-sequencing technologies, scientists have examined the patterns of mutations in a variety of cancer types to understand the factors involved. Different processes leave a different "mutational signature" depending on the exogenous or endogenous DNA damaging agents, as well as the repair or replicative pathways that follow (Helleday et al., 2014). Perhaps there are other places where endogenous enzymes are playing an important functional role as they alter the DNA sequence. If so, these signatures identified in cancer studies could help identify those places. It may be that the presence of AID in oocytes, spermatocytes, and embryonic cells is related to the induction of adaptive germ-line mutations.

AID: A Role in Adaptation?

Detection of AID in primordial germ cells, embryonic stem cells, and several other cell types was the impetus for investigating a possible role of AID in demethylation (Matthews et al., 2014). While AID does appear to sometimes play a role in demethylation, it could play another role in these cells. It has been suggested that it plays a role in meiotic recombination. SPO11 is an important enzyme that initiates double-stranded breaks during meiotic recombination. In some assays, AID appears to partially rescue SPO11 deficiency (Barreto and Magor, 2011). However, no statistical difference was noted in the average recombination events between normal and AID null mice (Cortesao et al., 2013). A third possibility that has been suggested is that AID may still play an APOBEC3-like role in controlling transposable element movement in some species (Barreto and Magor, 2011).

Another possibility exists: AID may be purposefully recruited to germ cells for DNA editing. In other words, enzymes such as AID and/or mutagenic repair pathways may be involved in the purposeful formation of adaptive alleles. It has already been noted that homologous recombination (crossing over and gene conversion) is mutagenic and that this is associated with adaptive mutations in bacteria. Error-prone TLS polymerases and/or error-prone repair pathways have been shown to play a role (Malkova and Haber, 2012). Given the purposeful nature of mutations induced by these mechanisms in adaptive immunity, it is quite plausible that genetic adaptation has a similar underlying basis.

If many germ-line mutations are purposeful, then it is expected that various factors will eventually be identified that govern the targeting of sites for mutagenesis and recruiting of appropriate proteins. Physiologic adaptation is characterized by changes in gene expression, which is mediated by epigenetic changes. Adaptive alleles often arise in the same genes as those targeted in physiologic adaptation (Lightner, 2014). Since transcription and associated epigenetic changes are important in recruiting AID, it may be that these factors play a role in targeting various regions of the genome for adaptive genetic changes.

Further, there may be purposeful mechanisms to increase the frequency of adaptive alleles in the population. Meiotic drive refers to any process that distorts Mendelian inheritance by preferentially transmitting one haplotype (or allele) over another when gametes are formed by meiosis. Biased gene conversion is one example; it can result from the break being preferentially induced on one strand of DNA over the other. Other downstream factors, such as the factors recruited to repair the break, can be involved as well. In some cases, this biased transmission is associated with, and perhaps influenced by, single nucleotide polymorphisms (SNPs; Odenthal-Hesse et al., 2014).

The existence of meiotic drive has significant implications for the evolutionary assumption that natural selection is a major player in adaptation. It has long been known, based on mathematical models that natural selection cannot account for diversity in vertebrates, even in an evolutionary time frame (Haldane, 1922; Kimura, 1968). Further, mathematical modeling suggests that genetic drift eliminates the majority of rare beneficial alleles. Finally, the phenotypically based prospective study of natural selection in Galapagos finches showed natural selection acted only at discrete times of harsh environmental conditions and was not consistent in direction (Lightner, 2015).

Despite this, there have been many genetic studies that have relied on statistical tests that suggest that natural selection has occurred, even within groups that creationists would say belong to the same kind. Yet these tests usually assume Mendelian inheritance, and the existence of meiotic drive violates those assumptions. Thus, meiotic drive is likely to be a well-regulated, designed mechanism (evolutionists currently assume it is random) that accounts for the statistical patterns normally attributed to natural selection (Lightner, 2015).

Summary

Historically, evolutionists have insisted that adaptation takes place by the naturalistic mechanisms of random genetic mutation and natural selection. These are philosophical assumptions based on a worldview that rejects a Designer, not something that was demonstrated scien-

tifically. In the creation model, universal common ancestry is rejected, but there is diversification and speciation that has obviously occurred within many created kinds. The appearance of adaptive alleles in various populations around the world suggests that there are designed mechanisms by which these alleles arise.

A look at the adaptive immune response indicates that the body has the ability to edit DNA in a variety of ways to meet environmental challenges. Several enzymes, including AID, are designed to initiate alterations in the DNA sequence. Then a variety of outcomes are possible depending on which proteins are recruited to resolve the aberrant base or DNA breaks that were induced. The whole process is well designed and tightly regulated. This bolsters the idea that similar designed mechanisms are involved in adaptive germ-line mutations.

Additionally, AID has been detected in ovaries, spermatocytes, and embryonic stem cells. While AID does appear to play a limited role in demethylation, which may at least partially explain its presence in these locations, this also leaves open the possibility that AID may play other roles. There is evidence it can partially cover for the loss of the enzyme SPO11, which is normally involved in meiotic recombination. AID, or other DNA editing enzymes, may be actively recruited during meiosis, which could help explain why homologous recombination is mutagenic beyond the normal crossing over and gene conversion.

Further, it is recognized that genes that are involved in physiologic adaptation are often the same genes involved in adaptive genetic mutations. Physiologic adaptation involves epigenetic changes that up- or down-regulate genes to compensate for an environmental challenge. AID can target regions where there is active transcription and other epigenetic signals, which suggests physiologic adaptation may plausibly be an important prerequisite if AID in fact does play a role in adaptive germ-line mutations. Since such mutations leave "signatures," it may be possible to bioinformatically screen for regions of the genome where adaptive mutations were induced by AID or a similar DNA editing enzyme.

Finally, when new adaptive alleles appear in a population, there needs to be an effective means for them to spread. Natural selection could play some role, but there are multiple lines of evidence suggesting that it is not particularly effective. The reality that meiotic drive, a type of non-Mendelian inheritance, exists suggests that it may play an important role in increasing the prevalence of adaptive alleles within a population.

All these considerations point to potentially fruitful lines of research. While the intelligent-design framework does recognize design is present in the genome, the history in Genesis provides background information suggesting several important places to look for this design. If AID and/or similar enzymes play a role in adaptive germ-line mutations, it would be one more line of evidence that adaptation occurs because of the Creator, the God described in the Bible, who cares for His creatures.

References

Arakawa, H., H. Saribasak, J.M. Buerstedde. 2004. Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. *PLoS Biology* 2:0967–0974.

Barreto, V.M., and B.G. Magor. 2011. Activation-induced cytidine deaminase structure and functions: a species comparative view. *Developmental and Comparative Immunology* 35:991–1007.

Belyayev, A. 2014. Bursts of transposable elements as an evolutionary driving force. *Journal of Evolutionary Biology* 27:2573–2584. www.ncbi.nlm.nih.gov/pubmed/25290698

Cantaert, T., J.N. Schickel, J.M. Bannock, Y.S. Ng, C. Massad, T. Oe, R. Wu, A. Lavoie, J.E. Walter, L.D. Notarangelo, W. Al-Herz, S.S. Kilic, H.D. Ochs, S. Nonoyama, A. Durandy, and E. Meffre. 2015. Activation-induced cytidine deaminase expression in human B cell precursors is essential for central B cell tolerance. *Immunity* 43:884–895.

Chandra, V., A. Bortnick, and C. Murre. 2015. AID targeting: old mysteries and new challenges. *Trends in Immunology* 36:527–535.

Cortesao, C.S., R.F. Freitas, and V.M. Barreto. 2013. Activation-induced cytidine deaminase does not impact murine meiotic recombination. *G3: Genes, Genomes, Genetics* 3:645–655.

Dominguez, P.M., and R. Shaknovich. 2014. Epigenetic function of activation-induced cytidine deaminse and its link to lymphomagenesis. *Frontiers in Immunology* 5:642.

Dong, J., R.A. Panchakshari, T. Zhang, Y. Zhang, J. Hu, S.A. Volpi, R.M. Meyers, Y.J. Ho, Z. Du, D.F. Robbiani, F. Meng, M. Gostissa, M.C. Nussenzweig, J.P. Manis, and F.W. Alt. 2015. Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching. *Nature* 525:134–139.

Haldane, J.B.S. 1922. Sex-ratio and unisexual sterility in hybrid animals. *Journal of Genetics* 12:101–109.

Harris, R.S., and J.P. Dudley. 2015. APOBECs and virus restriction. *Virology* 479–480:131–145.

Helleday, T., S. Eshtad, and S. Nik-Zainal. 2014. Mechanisms underlying mutational signatures in human cancers. *Nature Reviews: Genetics* 15:585–598.

Huxley, J. 2010. *Evolution: The Modern Synthesis*. The MIT Press, Cambridge, MA.

Jønsson, K.A., P.H. Fabre, S.A. Fritz. R.S. Etienne, R.E. Ricklefs, T.B. Jørgensen, J. Fjeldså, C. Rahbek, P.G.P. Ericson, F. Woog, E. Pasquet, and M. Irestedt. 2012. Ecological and evolutionary determinants for the adaptive radiation of the Madagascan vangas. *Proceedings of the National Academy of Sciences* 109:6620–6625. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3340096/

Jung, D., C. Giallourakis, R. Mostoslavsky,

and F.W. Alt. 2006. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annual Review of Immunology* 24:541–570.

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* 217:624–626.

Kohli, R.M., R.W. Maul, A.F. Guminski, R.L. McClure, K.S. Gajula, H. Saribasak, M.A. McMahon, R.F. Siliciano, P.J. Gearhart, and J.T. Stivers. 2010. Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *The Journal of Biological Chemistry* 285:40956–40964.

Kumar, R., L.J. DiMenna, J. Chaudhuri, and T. Evans. 2014. Biological function of activation-induced cytidine deaminase (AID). *Biomedical Journal* 37(5): 269–283.

Lamichhaney, S., J. Berglund, M.S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.J. Rubin, C. Wang, N. Zamani, B.R. Grant, P.R. Grant, M.T. Webster, and L. Andersson. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518:371–375.

Lanning, D.K., and K.L. Knight. 2015. Diversification of the primary antibody repertoire by AID-mediated gene conversion. *Results and Problems in Cell Differentiation* 57:279–293.

Le, Q., and N. Maizels. 2015. Cell cycle regulates nuclear stability of AID and determines the cellular response to AID. *PLoS Genetics* 11(9): e1005411.

Lerner, H.R.L., M. Meyer, H.F. James, M. Hofreiter, and R.C. Fleischer. 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian Honeycreepers. *Current Biology* 21:1838–1844. http://www.ncbi.nlm.nih.gov/pubmed/22018543

Lightner, J.K. 2006. Identification of species within the sheep-goat kind (Tsoan monobaramin). *Journal of Creation* 20(3): 61–65. http://creation.com/images/pdfs/tj/j20_3/j20_3_61–65.pdf

Lightner, J.K. 2008. Genetics of coat color I: the melanocortin 1 receptor (MC1R). *Answers Research Journal* 1:109–116.

Lightner, J.K. 2009a. Karyotypic and allelic diversity within the canid baramin (Canidae). *Journal of Creation* 23(1): 94–98. http://creation.com/images/pdfs/tj/j23_1/j23_1_94–98.pdf

Lightner, J.K. 2009b. Genetics of coat color II: the agouti signaling protein (*ASIP*) gene. *Answers Research Journal* 2:79–84.

Lightner, J.K. 2010a. Post-Flood mutation of the KIT gene and the rise of white coloration patterns. *Journal of Creation* 24(3): 67–72.

Lightner, J.K. 2010b. Identification of a large sparrow-finch monobaramin in perching birds (Aves: Passeriformes). *Journal of Creation* 24(3): 117–121. http://creation.com/sparrow-finch-baramin

Lightner, J.K. 2014. Adaptation of endotherms to high altitudes. *Creation Research Society Quarterly* 50:132–140.

Lightner, J.K. 2015. Natural selection: assessing the role it plays in our world. *Answers Research Journal* 8:111–119.

Malkova, A., and J.E. Haber. 2012. Mutations arising during repair of chromosome breaks. *Annual Review of Genetics* 46:455–473.

Matthews, A.J., S. Zheng, L.J. DiMenna, and J. Chaudhuri. 2014. Regulation of immunoglobulin class-switch recombination: choreography of noncoding transcription, targeted DNA deamination, and long-range DNA repair. *Advances in Immunology* 22:1–57.

Mayr, E. 2001. *What Evolution Is*. Basic Books, New York, NY.

Moris, A., S. Murray, and S. Cardinaud. 2014. AID and APOBECs span the gap between innate and adaptive immunity. *Frontiers in Microbiology* 5:534.

Nabel, C.S., E.K. Schutsky, R.M. Kohli. 2014. Molecular targeting of mutagenic AID and APOBEC deaminase. *Cell Cycle* 13(2): 171–172.

Odenthal-Hesse, L., I.L. Berg, A. Veselis, A.J. Jeffreys, and C.A. May. 2014. Transmission distortion affecting human noncrossover but not crossover recombination: a hidden source of meiotic drive. *PLoS Genetics* 10(2): e1004106.

Pettersen, H.S., A. Galashevskaya, B. Doseth, M.M.L. Sousa, A. Sarno, T. Visnes, P.A. Aas, N.B. Liabakk, G. Slupphaug, P. Sætrom, B. Kavli, and H. E. Krokan. 2014. AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature. *DNA Repair* 25:60–61.

Purdom, G., and K. Anderson. 2009. A creationist perspective of beneficial mutations in bacteria. Answers in Genesis: Answers in Depth. https://answersingenesis.org/genetics/mutations/a-creationist-perspective-of-beneficial-mutations-in-bacteria/ (accessed November 5, 2015).

Ramiro, A.R., and V.M. Barreto. 2015. Activation-induced cytidine deaminase and active cytidine demethylation. *Trends in Biochemical Sciences* 40(3): 172–181.

Reddy, S., A. Driskell, D.L. Rabosky, S.J. Hackett, and T.S. Schulenberg. 2012. Diversification and the adaptive radiation of the vangas of Madagascar. *The Royal Society Proceedings. Biology* 279:2062–2071. http://rspb.royalsocietypublishing.org/content/279/1735/2062.long

Rensch, B. 1980. Historical development of the present synthetic neo-Darwinism in Germany. In Mayr, E., and W.B. Provine (editors), *The Evolutionary Synthesis: Perspectives on the Unification of Biology*, pp. 284–303. Harvard University Press, Cambridge, MA.

Saugar, I., M.Á. Ortiz-Bazán, and J.A. Tercero. 2014. Tolerating DNA damage during eukaryotic chromosome replication. *Experimental Cell Research* 329:170–177.

Shan, E.L. 2009. Transposon amplification in rapid intrabaraminic diversification. *Journal of Creation* 23(2): 110–117.

Terborg, P. 2008. Evidence for the design of life: part 2—Baranomes. *Journal of Creation* 22(3): 68–76.

Terborg, P. 2009a, The design of life: part 3—an introduction to variation-inducing genetic elements. *Journal of Creation* 23(1): 99–106.

Terborg, P. 2009b. The design of life: part 4—variation-inducing genetic elements

and their function. *Journal of Creation* 23(1): 107–114.

Wayne, R.K., and B.M. von Holdt. 2012. Evolutionary genomics of dog domestication. *Mammalian Genome* 23:3–18. http://www.ncbi.nlm.nih.gov/pubmed/22270221.

Wood, T.C. 2002. The AGEing process: rapid post-Flood intrabaraminic diversification caused by altruistic genetic elements (AGES). *Origins* 54:5–34.

Yang, W. 2014. An overview of Y-family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry* 53(17): 2793–2803.

Zan, H., and P. Casali. 2013. Regulation of *Aicda* expression and AID activity. *Autoimmunity* 46(2): 83–101.

Creation Research Society Quarterly 2016. 52:275–308.

Cells as Information Processors

Part I: Formal Software Principles

Royal Truman*

Abstract

Cells perform millions of Boolean logic operations every second using multiple independent codes with stringent formal rules instantiated on DNA, RNA, proteins, sugars, and membranes. These codes rely on elementary and concatenated symbols to define variables and values that can be written, deleted, and read from long- and short-term memory. Computer and cellular variables are used with control structures such as "GoTo," subroutine calls, "wait," and to initiate and terminate iteration loops. They have well-defined data types and allowed operations. Values can be structured in arrays and linked lists.

Although variables are identifiable in cells, logic is executed without a readable source code, using hardwired biochemical components and inherited molecular machines (MMs). Each code requires unique decoding MMs, and cellular codes interoperate to incorporate details located throughout the cell to permit holistic correct decisions. Tight integration between these codes is implemented using adaptor biomolecules. DNA, RNA, and proteins are used to define both variables and values for independent codes, often in overlapping regions. These biomolecules are also needed to create MMs, adaptors, and the rest of the infrastructure.

Introduction

Biological research and interpretation have been dominated by philosophical naturalism for almost two centuries, especially when considering the question of origins. Deliberate design is often rejected as unscientific, which leads to even absurd proposals being entertained since "something must have happened." This is remarkable, since we interact daily with a world affected by conscious decision making. If we found a computerlike object on Mars, most would not insist on finding an explanation limited to deep time, random mutations, natural selection, chemistry, and physics. Although it would be possible to also explain the actions of a chess-playing program post-facto by tracing a series of internal mechanistic steps, this explanation would be incomplete. It would fail to explain the innate ability to anticipate and solve novel complex problems.

Prokaryote and eukaryote cells can do far more than a chess-playing program, being able to solve an aston-

* Dr. Royal Truman, Mannheim, Germany, royaltruman@yahoo.com
Accepted for publication April 22, 2016

ishing variety of unrelated problems concurrently. A seemingly endless list of contingencies has been anticipated, even when the exact details were never encountered before by the cell or its ancestors. Flexible categories of problems have been foreseen. Cells perform logic processing in a manner surprisingly similar to computers, using codes, structured datatypes, variables, algorithmic constructs such as Boolean logic and iteration, and a hierarchy of sophisticated data storage strategies for short- and long-term memory. Ignoring this integrated, holistic aspect of cells and insisting on a reductionist neo-Darwinian explanation for every cellular feature prevents answering the relevant questions correctly: Where did they come from, and why are they there?

Interpreting Biological Change and Development

Many complex processes exhibited by living systems suggest an intention or purpose. Examples include migration of birds to specific locations during certain time periods, development of adults from a fertilized cell, metamorphosis of caterpillars into butterflies, and execution of a strategy based on mental processes. This led philosophers long ago to embed purpose in physical objects as a form of internal will. Aristotle identified four kinds of causes for movement and change in general—the material, formal, efficient, and final—and claimed in Book II of *Physics* that a stone falls because it has an internal nature that drives it to attain its natural state. Many prominent thinkers since then have tried to interpret the specialness of living systems using notions such as a "formative drive," "living principle," "life-energy," "entelechy," and "teleology."

Currently, however, science has become dominated by reductionist and mechanistic thinking typified by books such as Jacques Loeb's *The Mechanistic Conception of Life* published in 1912 and the works of behaviorist psychologists—in particular B. F. Skinner—who deny the existence of will and mental states that perceive and direct behavior. This misguided naturalist thinking distorts much of what we observe and experience. Purpose and guidance are apparent and need to be taken into account. The existence and operation of an orchestra, growth of trees, poker-playing programs, and so on cannot be adequately explained by extrapolation from the natural behavior of many atoms. Wilhelm Dilthey (1833–1911), prominent philosophy professor at the University of Berlin, had a special interested in scientific methodology and introduced a distinction between the humanities (*Geisteswissenschaften*) and natural sciences (*Naturwissenschaften*). He argued correctly that investigative methods are often being applied in areas they are unsuitable for.

Purpose and guidance in nature need to be revisited. In this two-part series, we will examine how intent is governed in cellular processes, using digital computers as a model. We will show formal software principles are involved, which are processed by hardware molecular machines (Scruton, 1996, p. 254). University of Chicago microbiology professor James Shapiro referred to such stored instructions in a recent lecture, pointing out, "Cells use cognitive processes (=action based on knowledge) in dealing with genomic information" (Shapiro, 2011). At the conclusion of this analysis, we are reminded of Aristotle's claim that we cannot understand any cause for change until we can deduce its purpose (Stangroom and Garvey, 2005, p. 17).

Examples of Complex Programs in Cells

Prokaryote and eukaryote cells contain hundreds of integrated and carefully regulated programs such as metabolic networks and signal cascades linking the environment with gene regulation. Complex multicellular organisms display gene regulatory networks to unfold developmental programs and generate nervous systems and brain microcircuitries (Markram et al., 2015). We will examine these and other examples below and in the next paper. In all cases well-defined, logic-processing steps are involved, which channel the outcomes.

Coded Information Systems

In a series of papers, Truman introduced the theory of coded information systems (CISs), a framework to interpret how information-driven systems work. A CIS consists of linked tools or machines that refine outcomes to attain a specific goal (Truman, 2012a, 2012b, 2012c, 2013, 2015) (Figure 1). A coded message must play a prominent role between at least two members of these linked processes to demarcate from simple machines. Messages satisfy rules and strict formalisms to be interpreted reliably and provide flexibility and multipurposes (Hofstadter, 1980, p. 26).

Intended outcomes are ensured in a CIS through refinements using a combination of four possible "refinement factors": coded messages, sensors, physical hardware, and preexisting resources such as data or logic-processing algorithms. The model is quantitative, measuring the decreased entropy with respect to a reference state between each refinement step.

Often the CIS first increases the range of possible outcomes before applying constraining processes. To illustrate, the coding portion of a particular gene specifies a subset of useful protein sequences. How has entropy been decreased? The reference entropy to compare against is the variety of polypeptides that could be generated thanks to the cellular environment (without DNA, RNA polymerase, ribosome, ATP, tRNAs, and other resources, these long linear chains do not form naturally). The

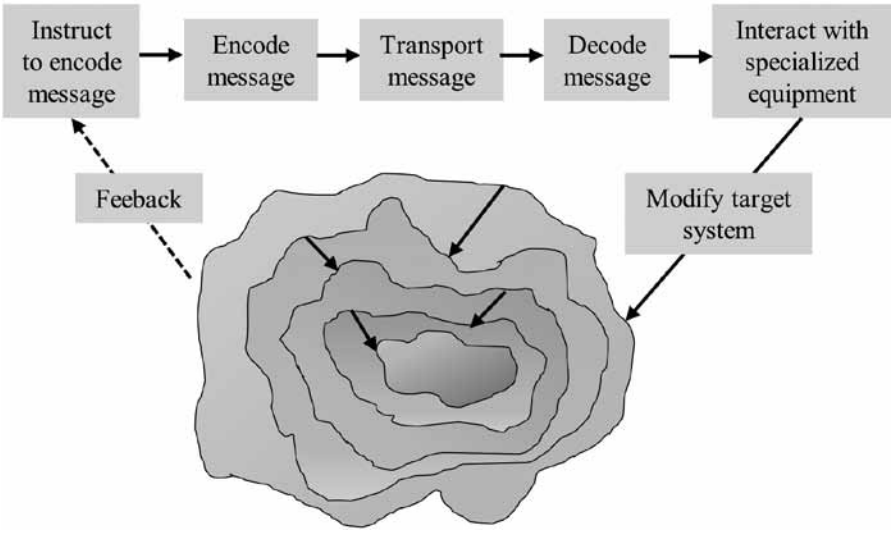


Figure 1. Coded Information Systems sequentially refine behavior through a series of processes. Each goal-directing refinement step could be influenced through coded messages, sensors, physical hardware, or preexisting resources such as data or logic-processing algorithms. At least one process must be guided by coded instructions to be a CIS.

reduction in the entropy of the reference sequences versus the sequences coded by a gene for a specific purpose defines the information gain.

CIS are often embedded hierarchically. The F_O region of ATP synthase is a component of the ATP synthase molecular machine, which is embedded in a mitochondrion, which is part of a cell, which is part of an organ, which itself is part of an integrated organism that contributes to an ecological CIS. Coded messages communicate intention between members of the system. In eukaryotes, many subsystems comprise an individual organism, whereas in prokaryotes there is more distribution of effort between collaborating species in an ecology with exchange of signals and genetic materials via passive uptake of DNA (Claverys et al., 2006), conjugal transfer, viral transduction, and other lateral gene transfer mechanisms (Stanton, 2007).

Indications Cells Could Be Computerlike

Modern computer architectures (Von Neumann architecture, n.d.) remind us of cells. DNA provides long-term storage, and the data are not randomly thrown together but sensibly structured, even as computers use file systems to organize related data. Genes in prokaryotes that need to be co-expressed are often located together and controlled by an operon (Osbourn and Field, 2009). In a recent study, for every eukaryote analyzed, gene order was not statistically random, but often those having similar and/or coordinated expression are clustered (Hurst et al., 2004; Michalak, 2008; Chu et al., 2011). Just as data on computer hard disks are stored in sectors, Alu-sequence containing nucleosomes define regions of the DNA (Salih et al., 2008; Trifonov, 2011).

DNA is a read/write/delete system. Data can be reorganized by transposons and content added via CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) (Zetsche et al., 2015; Ran et al., 2015), lateral gene transfer, and transfer of plasmids in prokaryotes. Genomes can also be contracted by deletions, such as the removal of transposable elements (van de Lagemaat et al., 2005). Portions of DNA are read many times and converted to mRNA copies—short-term memory—where logic processing is performed. Furthermore, mRNA codons specify amino acid sequences, so clearly a code exists.

We will focus here in Part 1 on formal software features like data types, data structures, codes, and algorithms, which are useful to solve problems using abstract methods, independent of the hardware implementation. The hardware aspects used by cells will be examined in Part 2.

Key Principles to Understand How Cells Work

Before showing that cells use formal software constructs, we need to devote some effort to eliminate a few misunderstandings and introduce some guiding insights: DNA does not provide an explicit prescriptive source program readable by humans; multiple codes are in use; each code requires a distinct alphabet and hardware decoder; software and hardware are far more integrated than in digital computers; and logic processing is distributed and hierarchical.

DNA Does Not Provide an Explicit Prescriptive Source Program

Many still erroneously believe DNA contains a prescriptive language containing a complete blueprint or "Book of Life:" that specifies in detail the development of an organism. As Woodward and Gills wrote recently, "This is the shock of shocks: that the DNA alone does not play the part of the director" (2012, p. 75). This contrasts with computer pro-

grams, whose logic can be understood from the source code. Consider as an example (1):

```
if (A=5 and B='red' and
not C='Deactivate')
then {'execute follow-
ing instructions'}    (1)
```

A line of readable coding such as (1) will not be found in DNA or elsewhere in a cell, but the variables can be identified, and logic operations are indeed being performed. Can we discern the Boolean logic and resulting processing being performed? Yes, empirically. Consider as an example of the variables A, B, and C three cis-regulatory elements (**CRE**s, specific nucleotide patterns on DNA). Each value is defined by which transcription factor (**TF**, a protein) is attached or "nothing is attached." The logic being performed can be deciphered by systematically varying the values (Davidson, 2006) and simulated with computer programs.

The logic is implicit but very real, and built into the system as whole, and for good reasons. Cells have far greater functionality than computers. They can replicate autonomously, generate their own energy, repair themselves, manufacture and recycle the substances needed, produce their processing hardware, and interact dynamically to provide emergent behaviors, even committing suicide (apoptosis) when necessary for the common good. An inheritable, error-free source code program to cover all these details and eventualities would not be feasible. Instead, cells replicate only the variables and their values, plus a functional copy of the necessary hardware each generation.

This strategy provides less opportunity for information corruption compared to specifying all the steps in precise detail in order to assemble thousands of cellular components, test the timing of location and progress of activity, and then mandate corrective action to take. We complete the explanation in Part 2 by showing how judicious organization—and inheritance—of the hardware components provide informative contributions and thereby reduce what the software needs to communicate.

Francis Crick was wrong when he claimed the genome was the (sole) source of phenotypic information (Crick, 1970). We can show this in many ways. A consequence of RNA editing, trans-splicing, and other post-transcriptional RNA modifications is that the modified sequences can undergo reverse-transcription and be introduced into the DNA germ line (Moller-Krull et al., 2008). Furthermore, changes in chromatin (which do not alter DNA sequences) can be inherited later over multiple generations (Jaenisch and Bird, 2003). In fact, somatically heritable chromatin structures are one way to establish differentiated cell lines (Gurdon et al., 2003). Further evidence that DNA does not directly prescribe final outcomes includes the existence of multiple life stages such as invertebrates having distinct larval and adult stages and other examples of metamorphosis. In the next paper, which accompanies this one, we describe the cell as an interacting set of controlling subsystems, each with its own coded variables, and less as a hierarchical or cascading design.

Much of what is necessary in the cell is not directly guided by DNA (Barbieri, 2003, p. 31). Globular proteins work only after they fold properly, which is affected by factors such as fluidity of the environment, how fast different sections are translated in a ribosome (Spencer et al., 2012), and the contribution of chaperones. Even after proteins form, additional guidance is provided, not by DNA, but by ligands, which are judiciously attached and removed. Gene regulatory networks develop automatically upon activating/deactivating CREs that are passively poised in anticipation. If one or more TFs activate a particular CRE, the resulting protein (a new TF) can activate or deactivate the same or different CRE(s), eventually leading automatically to mutually interacting circuits with no a priori guidance from explicit coded instructions.

RNAs can also behave as informative riboswitches. A small molecule binds to part of the RNA (the aptamer), which causes an allosteric change in another portion of the RNA called the expression platform, which can regulate gene expression (Serganov and Patel, 2007). There are many more examples of information processing that do not involve exclusive and direct guidance by DNA, such as aggregation of surface receptors in response to ligands (Wulfing et al., 2002; Bray and Duke, 2004; Murai and Pasquale, 2004) and cytoskeletal reorganization (Pollard and Borisy, 2003; Pelkmans, 2005).

There are cases, or course, where outcomes are partially specified directly by DNA, such as the N-end rule, whereby the half-life of proteins is determined to a large extent by the identity of its N-terminal residue. Sometimes DNA provides parameters less obviously such as in protein and vesicle targeting to distinct cellular locations (Bonifacino and Glick, 2004; Pool, 2005) and protein export from cells (Neel et al., 2005; Stuart and Ezekowitz, 2005). Here the signal sequences are extremely variable, both in length and amino acid composition, and the parameters are generated sometimes by remote parts of proteins brought together only after folding. This variability could be necessary for various processing details including additional post-targeting functions (Hegde and Bernstein, 2006; Emanuelsson, 2002; http://psort.hgc.jp/).

Evolutionists have generally argued that mutations are all that is needed to explain current cells. Distinguished Oxford professor Denis Noble, a forceful critic of Dawkins' reductionist views, pointed out that this is too simplistic: "Neo-Darwinism also privileges 'genes' in causation, whereas in multi-way networks of interactions there can be

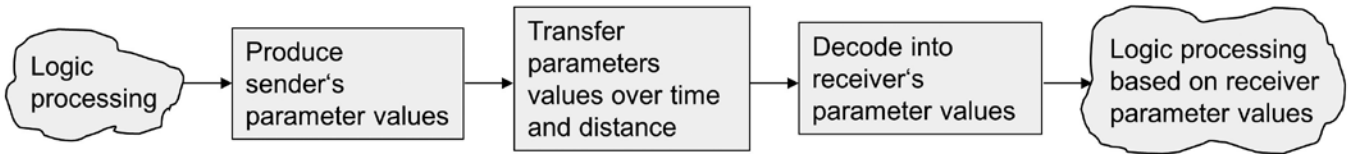


Figure 2. Logic processing can occur by the sender before communicating coded data and after the receiver knows what should be done. Sometimes little or no reasoning is needed to generate the sender's data, such as photons landing on retinas, and thereafter complex logic must be executed by the receiver to extract benefit from the data and decide what is to be done.

no privileged cause" (Noble, 2015 p. 1).

Does DNA determine outcomes by already possessing the necessary instructions, or does it respond to signals from the cell (e.g., to replace proteins decided by the cell are needed)? We agree with Noble, who also wrote, "The causality is circular, acting both ways: passive causality by DNA sequences acting as otherwise inert templates, and active causality by the functional networks of interactions that determine how the genome is activated" (2015, p. 9) and that "IF-THEN-ELSE" type instructions are found in cells (p. 10).

An interesting consideration is where most of the decision making occurs in computers and cells (Figure 2). This issue arises in all sender-receiver forms of communication. In some cases, a message could provide very detailed instructions, and in other cases the message is (explicitly) minimally informative. When only variable values are communicated, sometimes the sender performs considerable logical preprocessing and then only provides what is relevant (which the receiver can easily process). In other cases, raw data are made available, and the receiver is responsible to make sense out of it.

In the first example, we will consider, the *sender* has performed most the important logic processing before sending the following coded data (2):

```
(Co='IBM'; Nr_stocks_
to_buy=510; When_to_
buy='16 o'clock CET')
```
(2)

The receiver now knows what to do (which stocks to buy, how many, and when). Considerable decision making occurs in cells in the sender environment before the concentration and location of TFs are specified, and the results are communicated and processed as variable values by the relevant CREs variables at the receiving side.

In the next example, the *receiver* must perform much deductive processing, since variable values are communicated whose significance need to be interpreted and evaluated (3):

```
(Co='IBM'; Stock_
change_in_price=0.1;
Weather='cloudy';
Winner='Manchester
United')
```
(3)

The receiver must now determine what is relevant and how it correlates with the decisions to be made. Human minds typically process raw data considerably before making a decision.

What Is a Code?

What is a code, and how does it relates to logic processing using variables? A code defines rules that translate physical or mental details—such as sounds, images, pressure, size, quantity, intention, or even a different code—between two independent systems using an agreed-upon abstract convention of symbols. Speaking and writing are examples, bridging gaps in location and time. A simple causal outcome based on only a mechanical effect does not use a code, so an axe blow does not split a log in two thanks to a code that communicates intention. Whether to swing an axe could be communicated, however, using an arbitrary symbol convention such as {thumbs up /thumbs down}, {0 / 1} or {oui / non}.

The sender and receiver can share the same symbol set (alphabet), like the International Flag Code for merchant ships and the use of 'Co' in (2) and (3) above. An example in cells is when a specific TF (sender value) interacts directly with a CRE (generating a receiver value). Another example is when a DNA template is used to generate DNA copies. The next nucleotide value to add to the growing chain is communicated directly.

Alternatively, the sender and receiver could use different alphabets and variables as long as there is an unambiguous way to map the symbol sets. In (2) above, the receiver could assign the value for "Co" to its own variable "Company" and also convert the time 16 o'clock according to its own time zone. This kind of linkage may require **adaptor molecules**

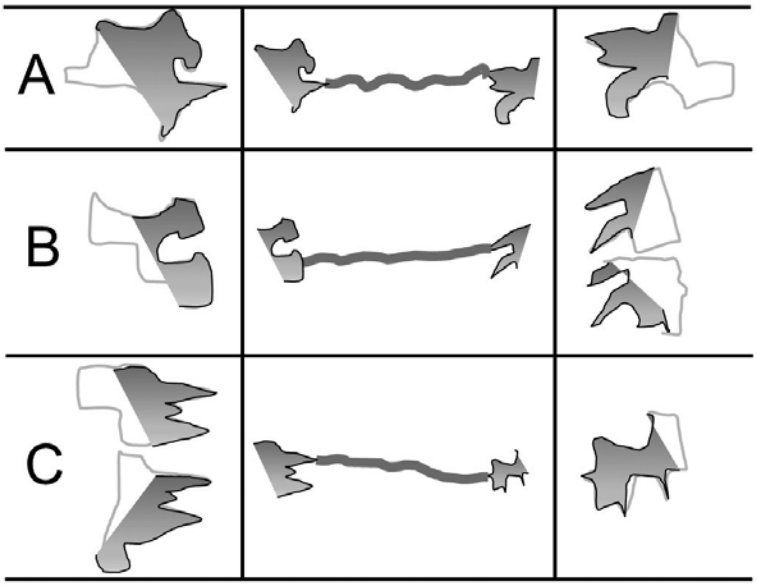


Figure 3. Communication between a sender and receiver system corresponds to transferring values to receiver variables. The alphabet of the sender (dark symbol in leftmost column) can differ from that of the receiver (dark symbol in rightmost column). In cells one or more adaptor molecule (middle column) may be needed to translate values between sender and receiver variables. The correct adaptor is identified through physical linkage with the sender variable's value.

A. 1:1 mapping between sender and receiver variable.
B. 1:n mapping.
C. n:1 mapping.

(Figure 3) or messenger molecules in cells.

If an informative ligand attaches to a TF, which then links to a CRE, this TF is now playing the role of an adaptor molecule. Another example of adaptor molecules are tRNAs in the genetic code, where one end identifies a specific mRNA codon (the value), and the other end translates to a corresponding receiver value (which activated amino acid to add to the growing protein). Linking the two systems through a chained network of signals permits additional factors to be taken into account that could refine the details during transfer.

Additional variables can be used within the sender and the receiver side to perform necessary reasoning. These can be independent codes, but at their interface there must always be pre-agreed conventions with respect to the meaning of the variables and how values are communicated. A receiver could then process the values assigned to its internal variables and then become a new sender, transmitting values to a new receiver. A chain of sender/receivers can result, and examples in cells include signal cascades.

Multiple Codes Are Used in Cells

Gordon Tomkins may have been the first scientist to propose that the genetic code is not the only code used in biology (Tomkins, 1975). Cell needs are communicated by different codes found on DNA, RNA, proteins, filaments, sugars, cell membranes, and other cellular components. Occasionally the literature seems to incorrectly claim a code is involved, such as the so-called **protein folding code** (Dill et al., 2008), in which multiple local activities occur in a precise order as part of the folding process. The difficulty in this case is identifying abstract variables upon which Boolean logic is performed. In this example, it seems that only physical chemical forces are occurring in a continuous and time-ordered set of steps. No variables are waiting to be assigned values nor anticipate activation.

Each code has its' own language and symbols. The **genetic code** to specify protein sequences is independent of the **DNA-binding protein code** to regulate gene expression (Hughes, 2008; Jolma et al., 2015) (Figure 4), even though both use DNA, and DNA codes sometimes share overlapping DNA nucleotides.

Entire collections of CREs can be organized into cis-regulatory modules (CRMs), leading to DNA code variants, since each CRM uses a separate set of rules. Figure 5 shows a representative example, where three exon are regulated by five such CRMs (Davidson, 2006, p. 49). Depending on time (e.g., development stage), input signals, and cell lineage, different modules can be used to interact with the key "proximal module" nearest to the transcription apparatus. This is a clear example of Boolean logic being applied.

In addition, by using different reading frames, the same code sometimes provides different messages. This was examined in mathematical detail for the genetic code at a recent conference on biological information (Montañez et al., 2013, pp. 139–167). In a remarkably candid paper, we read that "although dual coding is nearly impossible by chance, a number of human transcripts contain overlapping coding regions" (Chung et

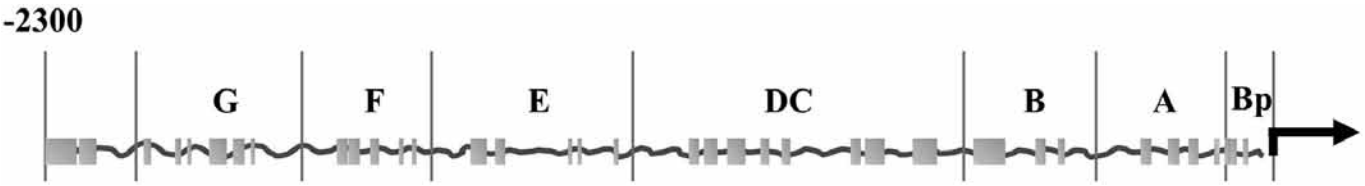


Figure 4. Representative example of cis-regulatory logic, showing the 2300 base-pair region preceding the coding region of gene endo16 of sea urchin. One or more proteins can bind to each of the cis-regulatory elements (gray boxes). The letters identify regions used for different purposes, such as regulation of key tissues during different phases of development (Davidson, 2006, p. 49–51).

al., 2007). These multiple codes prompted Trifonov to point out, "The times of surrender to 'junk' and 'selfish DNA' are over, and 'non-coding' becomes a misnomer" (Trifonov, 2011, p. 2).

We will not attempt an exhaustive listing of all cellular codes at this time, and the DECODE program continues to bring new ones to light, but we will mention a few to demonstrate that cellular codes define variables and their values but not procedural code as humanly readable instructions.

There is a **tRNA charging code** without which the genetic code cannot be implemented (Hou and Schimmel, 1988; Trifonov, 2011).

The **histone code** (Young, 2001; Jenuwein and Allis, 2001; Strahl and Allis, 2000; Cosgrove and Wolberger, 2005) involves post-translational modifications such as ubiquitination, phosphorylation, mono-, di-, tri-methylation, acetylation, sumoylation, and biotinylation of various residues on the four histones proteins (H2A, H2B, H3, and H4) that form the nucleosome. These tags regulate gene expression and other processes. Specific histone modifications can identify the need for DNA mismatch repair, for example H3K36me3 (histone H3, lycine number 36 receives three methyl groups) (Schmidt and Jackson, 2013) and H3K56 acetylation (Kadyrova et al., 2013). Hypoacetylation of H3K56 by enzymes HDACs 1 and 2 facilitate recruitment of nonhomologous end-joining (NHEJ) proteins (Miller et al., 2010). One should not overlook that each cell type in eukaryotes uses its own histone code (Carey, 2012, p. 188).

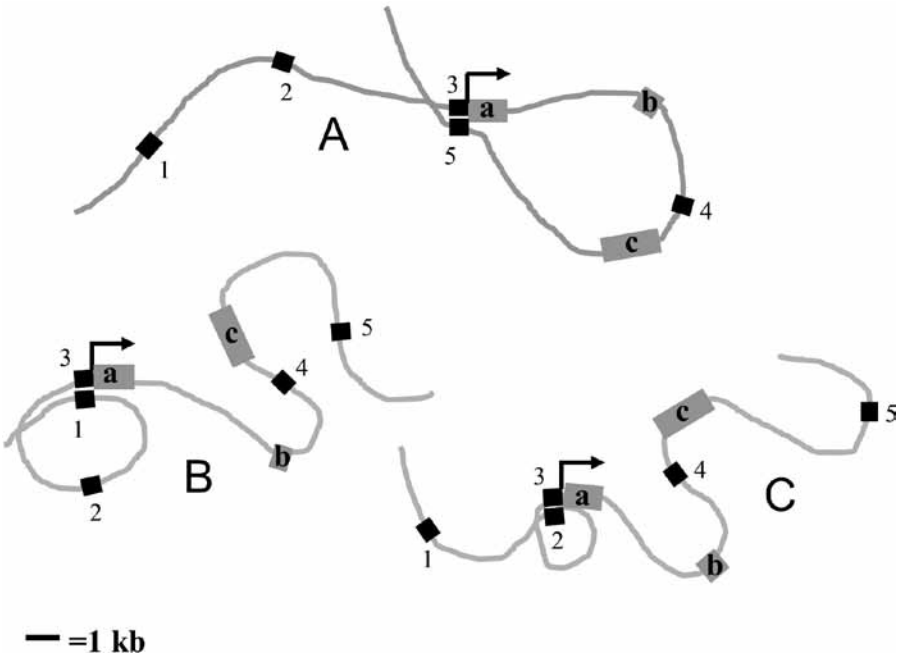


Figure 5. Multiple cis-regulator modules (CRMs) per gene, each composed of several CREs, permit independent regulation according to time, input signals and cell lineage. This typical example shows three exons (gray-checked boxes) regulated by five CRMs (black boxes). The CRMs are about 400 bp long, and the gene plus regulatory regions are spread out over about 30 kb of DNA. Alternative looping brings the relevant regions together (Davidson, 2006, p. 49). A: The "proximal module" 3 interacts with CRM 5; in B it interacts with CRM 1, and in C it interacts with CRM 2.

DNA methylation at the correct location identifies which sections of

DNA should be transcriptionally active euchromatin or inactive heterochromatin (Bird, 2002).

The **tubulin code** involves various ligands that are added and removed to microtubules to affect several cellular processes (Verhey and Gaertig, 2007; Janke, 2014).

The **splicing code** of eukaryote pre-mRNAs permits different exons to be combined to produce alternative proteins (Tejedor and Valcárcel, 2010).

The **nucleosome positioning codes**, also called "Chambon rules" (Barash et al., 2010), are understood well enough to algorithmically automate their location to within one base for biological DNA sequences (Segal et al., 2006; Trifonov, 1980; Trifonov, 1981; Gabdank et al., 2010). During development eukaryote genes are activated in a timed based manner using these codes for each primary transcript (Segal et al., 2006) to establish a regulatory circuitry that controls which genes are activated or silenced (Yuh et al. 1998).

Interaction between genes has also revealed the **Hox Code**. Just a few Hox or homeotic genes control development of the body plan along the anterior-posterior axis. They code for transcription factors, which can either activate or repress large gene networks. The same transcription factor can repress one gene and activate a different one, and TFs are involved at many levels within developmental processes (Wellik, 2007). A typical regulatory region in eukaryote DNA is about 500 nucleotides long, on which four or five transcription factors can bind. On average eukaryote genes seem to have about three such regulatory regions (Bray, 2009, p. 191).

The **N-end code** regulates the half-life of a protein using the identity of its N-terminal residue, which is determined from the moment they are produced (Varshavsky, 2011; Gibbs et al., 2014).

In the **sugar code**, oligomers of carbohydrates serve as ligands for the transfer of information, acting with lectin protein receptors (Gabius et al., 2011; Murphy et al. 2013). The large number of hydroxyl groups available offers enormous storage capacity, vastly more than the genetic code could (André et al., 2015).

The **adhesive code** (Readies and Takeichi, 1996; Shapiro and Colman, 1999) uses differences in adhesiveness between neural cells in the primordial neuroepithelium to first establish segmentation and then the emergence of specialized structures such as brain nuclei, cortical layers, fiber tracts, and neural circuits using cadherins.

A **niche code** has been proposed (Forsberg and Smith-Berdan, 2009). Hematopoietic stem cells (HSCs) must generate daughter HSCs and a variety of mature cells in response to stress in a regulated manner. HSCs are found in specialized niches in bone marrow, and there is a regulated adhesive interaction between niche cells and HSC components such as integrin, another example of adaptor molecules.

Signal Transduction Codes are used when extracellular signals ("first messengers" such as hormones, neurotransmitters, and paracrine/autocrine agents) attach to a specific receptor on the cell membrane, activating a smaller number of second messengers such as calcium, cAMP, nitric oxice, and phosphorylation cascades (Figure 6). One signaling molecule can cause many responses such as the cell's metabolism or gene expression, an example of 1:n variable mapping mentioned in Figure 3).

There is a vast research literature on this topic, and resources on signal transduction pathways are available on-line in databases such as "NetPath" for humans (http://www.netpath.org/). The latest research is correcting the view that simple linear cascades are used. Instead, large networks consisting of hundreds or thousands of proteins are involved (Walhout et al., 2013, p. 93). Note the rich potential to interact with other networks and codes to dynamically integrate multiple cell inputs and needs.

The actin cytoskeleton uses adapter molecules to identify materials that should interact there, which implies a **cytoskeleton code** (Barbieri, 2003; Barbieri, 2008, chapter 8).

The complex firing of neurons in the brain uses some kind of **neural code** or codes, since meaning is gleaned that permits the internal and external world to be understood (Nicolelis and Ribeiro, 2006; Cessac et al., 2010; Jessell, 2000; Marquardt and Pfaff, 2001; Flames et al., 2007). In spite of intense interest, it is far from being understood.

A **phosphorylation code** in Hedgehog signal transduction has also been identified (Chen and Jiang, 2013; Ficz, 2015; Schübeler, 2015).

The **miRNA code** can up or down regulate individual mRNA levels according to eukaryote cell type (Carey, 2012, pp. 191–194).

A **CpG epigenetic code** in eukaryotes governs millions of methylations on DNA. When near the gene start site, transcription is blocked but in the gene itself enhances expression (Jones, 2012). In this read/write/delete system, DNA-methyltransferases (DNMT) add methyl groups, and there are many mechanisms to remove them in a tissue-specific manner. Methylation is most dramatic in the brain (Keverne et al., 2015). Most of the methyl groups are removed in the fertilized egg (zygote) (Lee et al., 2014), otherwise the next generation would begin with a specialized and not pluripotent cell.

The **ventral neural tube** is an example of special codes used in cells that interpret a gradient concentration. Distinct classes of neurons are produced in the ventral neural tube according to local concentration of Sonic Hedgehog (Shh) (Briscoe et al., 2000).

Many secreted and membrane proteins contain N-terminal **signal sequences** that communicate their target locations (Hegde and Bernstein, 2006).

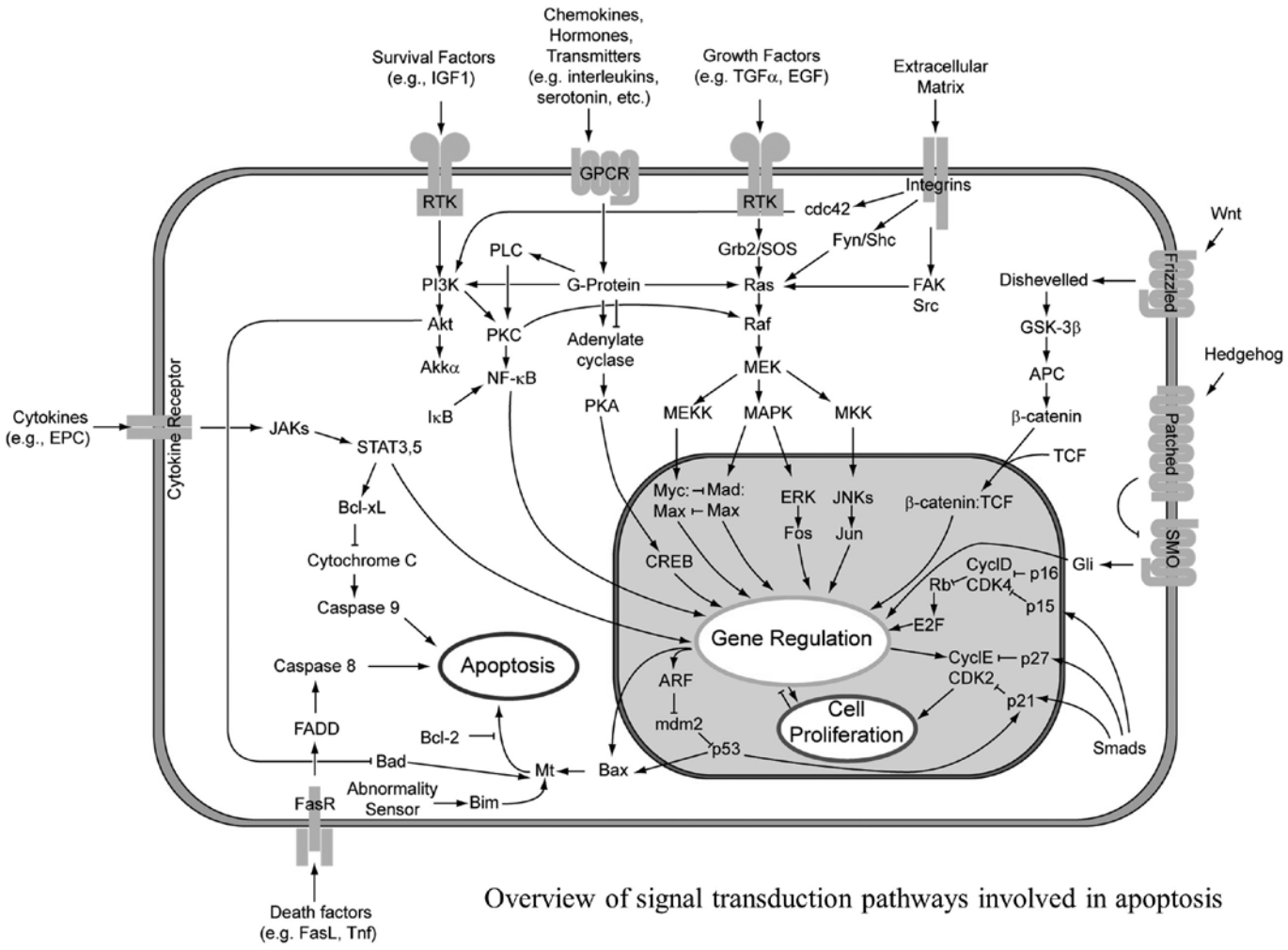


Figure 6. Example of a signal cascade pathway, here involved in programmed cell death (apoptosis). (Source of diagram: Wikimedia Commons, the free media repository, https://commons.wikimedia.org/wiki/File:Signal_transduction_v1.png). See also Klipp et al., 2009, pp. 135–142.

Codes in Cells Can Overlap

Cellular codes often overlap and therefore require degeneracy to not overly restrict each other. Since codes can be implemented using biochemicals which themselves rely on the genetic code, complex design tradeoffs are necessary. When planned correctly, the best implementation must be as robust as possible, taking into account the severity of possible errors for all the affected codes (mutations, mistranslation, etc.).

Degeneracy with respect to one code could be critically important for a different one. As an example, different codons could represent the same amino acid in the genetic code, but each codon can specify how rapidly that position is translated. Figure 7 describes this using a section of Java programming.

In probably all cases, assuming complete degeneracy for a code would be a mistake. Variants of a class of CRE could all be recognized by the same TF, but the CRE sequence differences specify how long and often to remain attached, in which tissue type, the timing of activity during a cell cycle, and for what stages of development.

The use of multiple and overlapping codes saves material and energy but is too constraining and requires too much foresight to find applicability in general purpose programming by humans.

```
public class TranslateCodons {
  public void ProcessEachCodon(String codon){
    String c = codon;
    switch (c) {
      case "GCU": {
        // 1: Delay translation by time t1. E.g., try {Thread.sleep(5);} catch (Exception e) {}
        // 2: Add Alanine to the growing chain
        break;
      }
      case "CGU": {
        // 1: Delay translation by time t2. E.g., try {Thread.sleep(10);} catch (Exception e) {}
        // 2: Add Arginine to the growing chain
        break;
      }
    // Remaining case statements…
    }
  }
}
```

Figure 7. Java example of codons being used for two unrelated purposes: to determine amino acid sequence and translation rate at that position of the mRNA.

Each Code Uses Its Own Processor

It is important to understand the distinction between variables and the values they can assume. Cellular variables possess recognizable steric and electronic features and wait for activation by a sender (which provides the values). For example, transcription in bacteria through RNA polymerase involves variables, like the "sigma factor recognizing promoter" (the -35 and -10 elements located before the beginning of the sequence to be transcribed). As possible values these locations could be unbound or bound to one of several possible "sigma factors." The sigma factor can also interact with a distinct set of promoters (Ishihama, 2000).

For each coding system there are special processors designed to interpret the relevant values. When TFs bind to cis-elements to regulate translation, an appropriate three-dimensional processor involving many proteins must be organized which can include direct or indirect adaptors (Zhou et al., 2015). The hardware aspect of cellular design is discussed in Part 2.

Software and Hardware Tightly Integrated

Unlike a Turing or von Neumann Machine (Von Neumann architecture, n.d.), cells must repair themselves, generate their own energy, adapt to new challenges, and reproduce autonomously with all necessary components over many generations. The solution is a complete synergistic interaction between the software and hardware. The physical DNA, RNA, and protein-based components that produce the hardwired biochemical processes are themselves constructed and replaced by relying on data provided through preexisting DNA, RNA and proteins.

It is often easy to identify the physical components of cells but overlook informational aspects. Each 260 million photoreceptors on a human retina could be identified, but the semantic content implied by the photons landing on them is then funneled on to only 2 million connected ganglion cells before sending to the correct processing regions of the central nervous system (Gazzaniga et al., 2009). Here information is being interpreted, compressed, and transferred.

As a second example, microtubules do much more than only maintain a

cell's shape. Per microtubule a hundred thousand or more globular protein units grow in many directions and degrade constantly until coming into contact with a specialized region of a chromosome centromere (Sullivan et al., 2001), or membrane, after a signal arrives there, at which point a firm attachment prevents degradation (Kirschner and Gerhart, 2005, pp. 148–152). These attachment regions are sensors (variables) that assume a value (i.e., when activated by the tip of the microtubule), that recruits proteins to produce a decoding molecular machine.

Logic Processing Is Distributed and Hierarchical

Different prokaryote species form ecological systems with necessary genes distributed among the members (Sonea and Mathieu, 2001), which is why a particular function requiring several genes can be assembled in one member through horizontal DNA transfer. Plasmids in prokaryotes are another example of **distributed** information processing. In eukaryotes, information processing is also distributed, such as when bacteria digest food separately from the host organisms' germ line. Different cell lineages also distribute the effort, where each cell type has characteristic ensembles of activated and deactivated genes. Proteins, polysaccharides, lipids, and other substances are used to interact with receptors on cell surfaces and provide communication signals to convey metabolic and developmental status back and forth (Aricescu and Jones, 2007; Takada et al., 2007; Yamada and Nelson, 2007; Widelitz, 2005). Intercellular communication also occurs by molecular diffusion through air or water using gases, amino acids, oligopeptides and vitamins as signals (Bogdan, 2001; Chen et al., 2005; Fuqua et al., 2001; Chambon, 1995; Lazazzera, 2001).

Hierarchical information processing also occurs. As examples, low-level logic processing occurs when individual DNA nucleotides define individual RNA nucleotides, and when codons specify amino acids. Once a protein has formed, additional processing occurs to transfer it to the correct cell location, later to integrate into molecular machines, enzymatic networks, and metabolic networks. Thereafter ever more complex features can develop, such as entire eukaryote organelles which themselves become part of a properly regulated cell, on up to organs, whose operations must also be carefully regulated to permit a viable organism that interacts within an ecology.

In addition to such hierarchical integration, we will see in the accompanying paper that many control systems in cells—each with their own codes—interact mutually within what often seems to be the same hierarchical level.

Generic Insights from Computer Systems Architecture

The explosive development of computer technologies is the result of collaboration between millions of scientists, engineers, and mathematicians worldwide. Fundamental to this success are interoperability conventions and standards (such as the Open Systems Interconnection model). This permits specialists in various hardware and software areas to focus on and develop technologies from which integrated systems result. Using these design insights, we will interpret cellular behavior by examining software and hardware aspects individually and consider different levels in the system at which guidance is provided.

Another insight humans have gained is the design of subsystems that can be assembled. We discussed lateral and hierarchical logic processing above (Thanbichler and Shapiro, 2008; Schneider and Grosschedl, 2007). An external printer can be built separately and then linked to the rest of the system. To work properly the hardware devices often also require their own dedicated software (e.g., "drivers" must be installed).

Software Elements Used to Implement Processing Logic

Before examining software constructs used by computers and cells, let us consider a simple program to calculate the factorial of a number (Figure 8).

Several general principles can be discerned.

1. The programmer did not need to consider how the solution would be implemented on hardware nor the operating system details. Only the logic needs to be accurately expressed symbolically.
2. There is a language with a precise grammar that contains several generic constructs—for example, iteration (with a defined starting and finishing value) and a *Boolean test* (if i has a value of n or less, then add 1 and continue, otherwise terminate the iteration).
3. The same processing logic could be applied with different values and meanings for the variable n.
4. The algorithm could be copied into other programs and modified.
5. The variables belong to a specific data type and have properties consistent with them. In the example, i and n must have an integer value: one cannot assign a value of "Smith" nor "True" to them.
6. The variables can represent real objects, like dollar bills, but the choice of the symbols and what they do are physically independent of what they specify.
7. The algorithm continues to make sense if each variable is replaced by another unique symbol. Even a three-dimensional abstract symbol could be used and the values assigned could also be represented by no code currently in use by computers. However, changes in hardware would then become necessary.

```
public class Factorial {
  public static void main(String[] args) {
    int n = 7;
    int result = 1;
    for (int i = 1; i <= n; i++) {
      result = result * i;
    }
  System.out.println("The factorial of " + n + " is " + result);
  }
}
```

Figure 8. Programming using Java to calculate the factorial of a number to illustrate the use of common software constructs to solve problems independent of the hardware implementation.

8. To have any value, the outcome from the algorithm needs to be retained or have some kind of effect.

All these and other principles can also be identified in cellular information processing. In the example in Figure 8, we see how limitless cases could be solved by merely replacing the numbers *i* and *n* as needed. This works only if programming constructs such as iteration, assignment of values to variables, and so on, exist. Otherwise a unique mechanical arrangement would be needed to solve each example. It is this use of general-purpose symbolic logic, which can be mapped to mental or physical objects, that is so special about computers and cells.

After this long, but necessary, preparation, we are finally ready to examine three important topics in the art of designing software: generic software data structures; generic programming elements; and file formatting. These are fundamental for computers.

I. Generic Software Data Structures.

Let us examine how data is usually structured in modern computers and cells to facilitate use in general-purpose programming constructs discussed in section II.

Symbols in an alphabet

Codes rely on an alphabet of elementary symbols. Modern digital computers use an alphabet of two symbols {0, 1} called bits. Cells use dozens of alphabets for their many codes. DNA is composed of four nucleotides abbreviated {A, C, G, T}, RNA also uses four nucleotides {A, C, G, U}, other codes rely on small ions such as cAMP (Ashcroft, 1997; Krysko et al., 2005) and calcium (Wagner et al., 2015), or on small parts of larger molecules.

One or several symbols taken jointly define an item, field, constant, variable, or value. In the past, telegraph messages used 5-letter commercial coded values such as BYOXO ("Are you trying to weasel out of our deal?") and LIOUY ("Why do you not answer my question?"). Other conventions also exist, such as LOL ("Laughing Out Loud") and CU ("See You"). In the extended ascii ISO 8859–1 code, '00001001' represents a *Line Feed*, '01000001' represents the letter *A*, and '00111000' represents the decimal digit 8. The codeword length of values can be fixed as in the asci extended and the genetic code or have different lengths as in compressed codes to store and transmit electronic data (Togneri and deSilva, 2003). There are design trade-offs to consider when deciding whether to use a fixed or variable length (Truman, 2012).

The symbols used by computer programs must be exact to be processed. C*onfirmation* and *Conformation* are almost identical, but not the same.

Different codes can be linked using different alphabets. A sender code could be restricted to a symbol from, e.g., {green, yellow, red}, which the receiver could translate to its system, e.g., limited to {1, 2, 3}.

When large molecules are used to convey coded meaning in cells, typically a small portion is informative, and the rest plays an adaptor molecule role or is used for the implementation details. Consider proteins. Portions of different residues are integrated to define a joint "symbol" having unique steric and electronic properties. The resulting symbols must be decoded using three-dimensional processors. In the fluid environment of cells under varying temperatures, the decoders must be more flexible than in computers. One consequence is that a portion of different amino acids could be combined to produce functionally the same symbol meaning in three dimensions.

Data types

Modern computer languages enforce data typing, which defines the kinds of values that can be assigned to variables to prevent errors. Common types include integer, floating-point number, character, alphanumeric string, and Boolean. Each kind of variable for biological codes is restricted to a range of values. The genetic code uses DNA and mRNA codons, whereas the enzyme complexes used by the histone code do not process codons (http://www.cellsignal.com/contents/resources-reference-tables/

histone-modification-table/science-tables-histone).

As another example, many mRNAs can interact with only some miRNAs (which specify what is to occur to the mRNA; Verdel et al., 2009; Sugiyama et al., 2005). This corresponds to 1:n, n:1, or n:m variable binding in Figure 3. In addition, only certain noncoding RNA data types (specific siRNAs, piRNAs, Alu RNAs etc.) are recognized by mRNA binding proteins.

Data type subsets

A subset of a data type can also be established for a specific program or module to further narrow acceptable values in some programming context—for example, only certain acceptable city codes for telephone numbers in a city, or a list of alphanumeric identifiers for a product line. We find this principle also in cells. The codons to represent Alanine must come from the subset {GCU, GCC, GCA, GCG} and Arginine from {CGU, CGC, CGA, CGG, AGA, AGG}.

Operation are defined for each data type

Specific computing methods or operations are permitted for each data type (and also for complex structures like matrices, arrays, etc.). One can negate a Boolean variable to convert True into False, but negating a data type "character" makes no sense. String variables can be concatenated, for example phrase = "white" + " " + "house" to form "white house," but this won't work for variables such as integers.

This principle is also found in cells. Each code used with DNA, RNA, proteins, sugars, or membranes is limited to its variable type(s) and their allowed operations. Consider the processing operations that can be performed with mRNA's data type "codon." The values can be read at the *A* (acceptor) or *E* (exit) portion of ribosomes (the receiver variables), they can be "concatenated" on each side to form polymers, and they can base pairs in unique ways (A-T and C-G). These kinds of operations cannot be assumed for other data types, such as hormones, transcription factors, or neurotransmitters. Ribonucleases and restriction enzymes can cut DNA strands using a subset of acceptable patterns (the receiver variable), but these locations are not processed on a codon basis as the genetic code does.

Group item

Elementary fields or items in computer languages can store values long-term using compound symbols. In many programming languages, several elementary items can also be combined and processed jointly for read and write purposes. As an example, a group item "*address*" could be composed of elementary items "*house-number*," "*street*," "*city*," and "*country-code*." Additional hierarchical clustering is also used in computer languages (such as C, Pascal, and Cobol), meaning group items can be further combined into records for example. This principle is also found in data transfer conventions like XML.

In cells, we recognize this principle whenever elements containing substructures are processed as a complete entity. One example is telomeres at the end of chromosomes, composed of groups of repetitive nucleotide patterns (e.g., TTAGGG in vertebrates), which are replenished by the enzyme telomerase reverse transcriptase. The six individual nucleotides are processed as an ensemble. In ***S. cerevisiae***, each $C_{1-3}A/TG_{1-3}$ repeat, taken jointly, constitute a potential binding site for Rap1 proteins, which recruit additional proteins (Williams et al, 2010).

In mammals, shelterin protein complexes regulate telomerase activity. Two of the six subunits (TRF1 and TRF1) bind uniquely to individual double-stranded TTAGGG (de Lange, 2010). So once again we recognize the concept of a grouping of elementary components. At a higher level, multiple copies of the individual patterns are treated as a new grouped entity and added all together to a chromosome by TERT (TElomerase Reverse Transcriptases) using a piece of template RNA known as TERC (Jády et al, 2006).

Group items consisting of smaller group items are not limited to repetitive patterns. Multiple codons are placed together within exons, which themselves are integrated into a primary RNA transcript. Processing as a whole occurs, such as in retrotranscription and rearrangements with the help of transposable elements.

The concept of group processing reminds us of how several residues jointly lead to discrete motifs in folded proteins and how a larger numbers of residues work together to form secondary structures such as alpha helices and beta sheets. Different nucleotide combinations also produce special RNA motifs.

Microbial genomes are also known to have an operon-like organization at various scalar levels (Audit and Ouzonis, 2003).

Concatenated index

In relational databases such as Oracle, a unique combination of one or more index values can be used to identify data records. Similarly, multiple nucleotides define promotors to identify the location of genes.

Array

Arrays and linked lists contain a series of values. In arrays, values of some datatype are stored in numerically indexed positions. The position within the array is informative and can be used directly in programming logic. If a certain value is always located at a specific index position (or a limited range of positions established in advance), it can be accessed directly by processing logic. An example using Fortran (a language well-suited to matrix calculations) is shown below (4). Assume that the results of a student's different exams are stored in known index

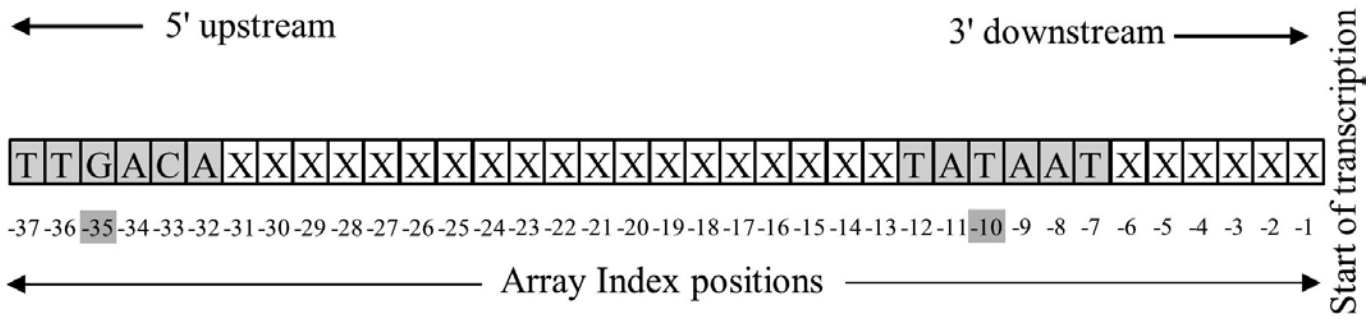


Figure 9. Nucleotide patterns at specific locations in bacteria define consensus promoter elements. The Pribnow box is centered at the -10, and a second component is often found at the -35 nucleotide position upstream from the start of transcription. Other regulatory elements are sometimes centered at the -41 or -61 position. If each nucleotide in the regulatory region is stored in an array, the index position can be used to program logical tests.

positions of array Examresults, and index position 3 contains the points obtained for the math test. The programing logic might look like this:

```
IF (Examresults(3) .GE.
70 .AND. Examresults(3)
.LT. 85) THEN Mathgrade
= 'B'                    (4)
```

Highly relevant to our discussion about cells, the value of interest could in principle be stored in different array positions if the acceptable alternatives are established in advance. Suppose there were two examiners and the result if determined by the first one is stored in Examresults(20) and if by the second examiner in Examresults(21). Now the program must determine the test results for the math exam by looking up the contents of array positions 20 and 21 and select the one having the exam result.

Prokaryote promotors illustrate array data storage and processing. For the Pribnow Box, a six-nucleotide consensus TATAAT is used by E. coli, centered at the -10 position, and often a second pattern TTGACA centered at -31 (Figure 9). For some bacteria or genes, the array positions to check could be slightly shifted, but legitimate indexed positions to be tested are known in advance. We will not elaborate here on the reasons for using alternative array positions, but it could be to regulate transcription rates or the results of genomic rearrangements.

There are many more examples of array processing in cells. In a typical ca. 22-nucleotide miRNA, usually only 6–8 adjacent or almost adjacent nucleotides (the seed region) at the 5' end are relevant, which is also true of the corresponding receiver variable on an mRNA. Logical tests on candidate miRNAs and their binding sites can therefore be performed using array index values. As another example, the coding parts of DNA and mRNA specify amino acid sequences, and the nucleotides need to be processed as triplets with no frameshifts. This permits translation to read the codons located in sequential index positions along mRNAs. In other words, each array index position does not contain a nucleotide, but a codon. Once the mature mRNA is ready for translation the length remains fixed, another characteristic of arrays.

Additional examples of processing array data include the symbols used by mobile elements to recognize insertion motifs; the portions of folded TFs that recognize cis-regulatory combinations; and the portions of enzymes that recognize restriction sites.

We see why many proteins must fold reliably into the same three-dimensional structure. This brings the relevant elementary symbols together so each can be assigned to a three-dimension index, "protein_position[i,j,k]." The relevant array positions refer to location in three-dimensional space and not the primary protein sequence. The resulting symbols need to be defined well enough to permit variables and their values to recognize each other, synergistically molding themselves together and avoiding false positives.

Whenever for a DNA or RNA-based code the distance between key nucleotide patterns are exactly or almost exactly known (including epigenetically modified nucleotides), then an indexed array seems to be a better description than a linked list. Knowing index values allows other array positions to be skipped and ignored. This is physically implemented in cells by constraining the decoders (e.g., portions of proteins) to specific

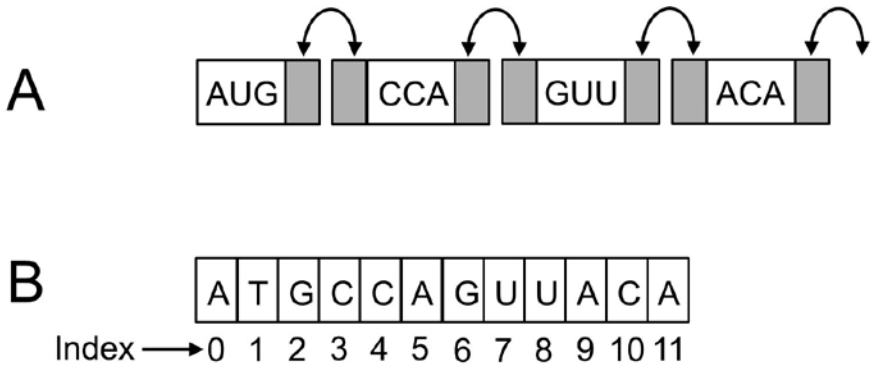


Figure 10. Linked lists and arrays. A. Double linked lists contain data (non-shaded boxes) and links (gray boxes), which point to the preceding and next member of the array. B. Arrays contain data at static locations identified by index values.

ranges of distance and location between the relevant data elements. (In the case of linked lists, however, a more complicated search for the relevant variables must be implemented).

We suggest below that DNA replication and transcription processing, which are used by different codes than those just discussed, are based not on arrays but linked lists. There are subtle differences between these kinds of data structures. For example, in computers the length of an array is established when the array is created (unlike linked lists, which grow and shrink as needed). Remarkably, in cells the same nucleotides are sometimes used by different codes concurrently, each with different kinds of data structures.

Linked List

A linked list is a chain of data and link values. The data part contains the useful information, and the link has the address of the next or previous element. Single-linked lists only point to the address of the next element, whereas double-linked lists include pointers to the next and the preceding data location (Figure 10).

Either an array or linked list could be used for programming purposes. They do differ, however, in internal implementation in ways that affect execution speed of data insertion, deletion, updating, and searching. One difference is that the index value where specific data is located in array lists is generally not known in advance and can change. Unlike arrays, linked lists can automatically grow and shrink dynamically as needed.

To illustrate the difference, candidate CRMs that could interact with the proximal module to regulate a gene are separated by distances that can vary (Figure 5). Finding the activated CRM requires a search for relevant data symbols whose positions are not defined by unique index values. An additional complication is that the regions of the CRM that are to bind to the proximal module involve CREs whose positions are not static in three dimension and must also be searched for.

The same reasoning applies when spliceosomes identify variable intron content whose boundary is defined by splicing signals (Rino and Carmo-Fonseca, 2009). The introns are generally not identifiable a priori by fixed index positions and the spliceosome succeeds even if transcription error adds or eliminates nucleotides.

In linked lists, elements of a defined data type (which could be a complex group of different item types) can be added to the end, inserted at any position, modified or removed (for arrays also, but that requires more processing effort). In addition, another linked list can be added on to another at any position. One disadvantage, of course, is that more effort is required to find a specific value compared to when its indexed location is known in advance.

In RNA, the four nucleotides {A, C, G, U} are attached to riboses (and deoxyriboses for DNA), which are held together along the backbone by phosphate groups (Figure 11).

Analogous to linked lists, nucleotides can easily be added, removed, or inserted simply by breaking and reattaching "address pointers," here phosphate bonds. This is an excellent description of what happens when DNA chains replicate one base after the other, RNA is transcribed, introns are removed, exons are spliced together, and chromosome crossover occurs. Absolute index values per se are generally not relevant for the logic processing, unlike for arrays.

We summarize in Table I some of the built-in methods available to linked lists, using the Java language (https://docs.oracle.com/javase/7/docs/api/java/util/ArrayList.html) and include some examples from cells.

In many cases, the processing could be defined in terms of linked lists and/or arrays. Let us recall miRNAs and take into account the concept of sublists, or relative indices, mentioned in Table I. In processing step 1, the nucleotides of a candidate miRNA could be assigned to a sequential linked list. In processing step 2, sliding windows 6–10 nucleotides long (representing candidate seed regions) could be fed into a fixed-length array. The values in array position[0] … position[9] would then be systematically tested against possible acceptor variables in mRNAs. Multiple hits are allowed.

Method	Meaning
Add()	Appends an element to the end or inserts at a specific position.
Cells: RNA transcription; some forms of RNA editing can insert codons (Bass, 2002; Nishikura, 2010); removing introns and splicing exons together; replicating DNA; chromosome cross-over.	
Clear()	Removes all of the elements from a list.
Cells: Upon degrading RNA all resources are free to be used for other purposes, unlike for arrays which when empty still consumes computer memory.	
Contains()	Returns true if this list contains the specified element.
Get()	Returns the element at the specified position in this list.
IndexOf()	Returns the index of the first occurrence of the specified element.
To identify introns, a primary transcript is searched to identify where it starts and ends to identify the index values. Intron lengths can vary considerably. Automated algorithms, such as SplicePort (http://spliceport.cbcb.umd.edu/) and GeneSplice (http://ccb.jhu.edu/software/genesplicer/) reflect the logic used in eukaryote cells to identify splice sites. The same concept is found in DNA in which transposable elements can be removed from genomes using patterns that define where they begin and end (van de Lagemaat, 2005). Other examples include: the initiation codon on mRNA is searched for (and modified) and so is the region on mRNA at which to create polyadenylation tails; patterns on mRNA are also searched for where nucleotide posttranscription modifications are to occur. The CRMs (Figure 5) are of variable distance from each other (e.g., after insertion of transposable elements into DNA) and need to be found. The location of elementary symbols for activated CRMs can also be variable, depending on what TFs are bound and which ligands these TFs contain.	

Table I. Some in-built methods used with linked lists in object oriented programming languages like Java and examples from cell biology.

Variables as a Data Structure?

We mentioned that in programming, arrays and linked lists are used to store **data values**. These can be assigned to a variable. For example, for an employee stored in index position 45 we might have a line of programming such as: SalaryInDollars = SalaryInPesos(45) * 1.4, and there is no ambiguity in how the value assignment occurs, nor in what was assigned to the variable "SalaryInDollars." Sometimes this is also true in cells. The anticodon of a specific tRNA is fixed, and the value of the communicated charged amino acid is exactly specified. But in cells this is not always that straightforward. It would be as if the variable SalaryInDollars could have small physical differences that affect how it interacts with the array positions, leading to significant effects. This issue can also apply to variable assignments that do not involve arrays and linked lists.

Unlike computers, cells often use variants of variables that do not respond identically to the same values. For example, a CRE is like a sensor, a variable that can be assigned values such as "TF_n bound" or "no TF is bound." However, the binding sequence of a particular CRE can vary and therefore respond differently to an identical TF (which itself can provide many values). This can have serious consequences, affecting how fast and long binding occurs, and could even affect the subsequent Boolean logic. (For example, a modified CRE might affect the geometry of the bound TF and thus how it interacts with other factors.)

This suggests a novel technical inspiration for computer scientists and bioinformatic researchers. Instead of

Method	Meaning
RemoveAll()	Removes from the list all occurrences of specific values.
Cells: Examples include tRNA splicing (Trotta et al., 1997) and RNA self-splicing (Cech, 2002) based on secondary or tertiary structure. These rely on discrete structures which can be stored as structured (i.e., multisymbol) values in individual linked list positions, which is a different operation than removing whatever is found between two boundary patterns. This assumes a specific code is to work with the linked list. Note: Gene silencing mechanisms are not the same as physical compacting through physical removal.	
RemoveRange()	Removes the elements whose index is between two specified indices.
Cells: After the index location of intron/exon boundaries are found, the introns can be removed from primary transcripts.	
Set()	Replaces the element at positions that need to be specified with a value.
Cells: Error correction mechanisms use a DNA or RNA template; any process which modifies a DNA nucleotide (like methylation) or RNA codon, including RNA editing (Bass, 2002; Nishikura, 2010).	
SubList()	Sublist data structures are a feature of linked lists and arrays. Logic processing is performed with respect to the sublist and its own indices, for which the first one is assigned an index value 0, the second 1, etc. All operations performed on the sublist are reflected in the original full list.
Cells: The seed region within miRNAs. In addition, many of the examples above rely on first identifying the location of boundaries; what is relevant thereafter are the relative positions.	

Table I (continued)

treating members of a class of CREs as functionally identical or as separate variables—as we have been implying so far—the suggestion here is to develop a fuzzy-logic type technology which permits both variables and values to be processed with variability. Finding cellular **variables** would then also use linked arrays, since the candidate regions and length would be unknown in advance. The imprecision of many bioinformatic software tools to identify regulatory patterns reflects these joint uncertainties.

Here is an example. RNA polymerase and TFs search for DNA (response elements, or sensors) in 100–1000 base-pair regions upstream from the transcription start site and on the same strand. Nucleotide positions are indexed with negative numbers counting back from -1 towards the 5' direction. The patterns to test are variables that are not always the same in location or details, which is where linked lists become useful. In focused initiation, transcription starts at a single nucleotide or within a narrow region of several nucleotides having sequence motifs such as the TATA box and DPE. In dispersed initiation, there are multiple weak start sites over a broad region of about 50 to 100 nucleotides (Juven-Gershon and Kadonaga, 2010).

This suggestion captures those cases where symbols seem to have both variable and value character. The regulatory portion of genes define variables that need data to know when and where to initiate transcription, but simultaneously RNA polymerase and TFs sometimes also provide variables that need data to know where to attach in the promoter region.

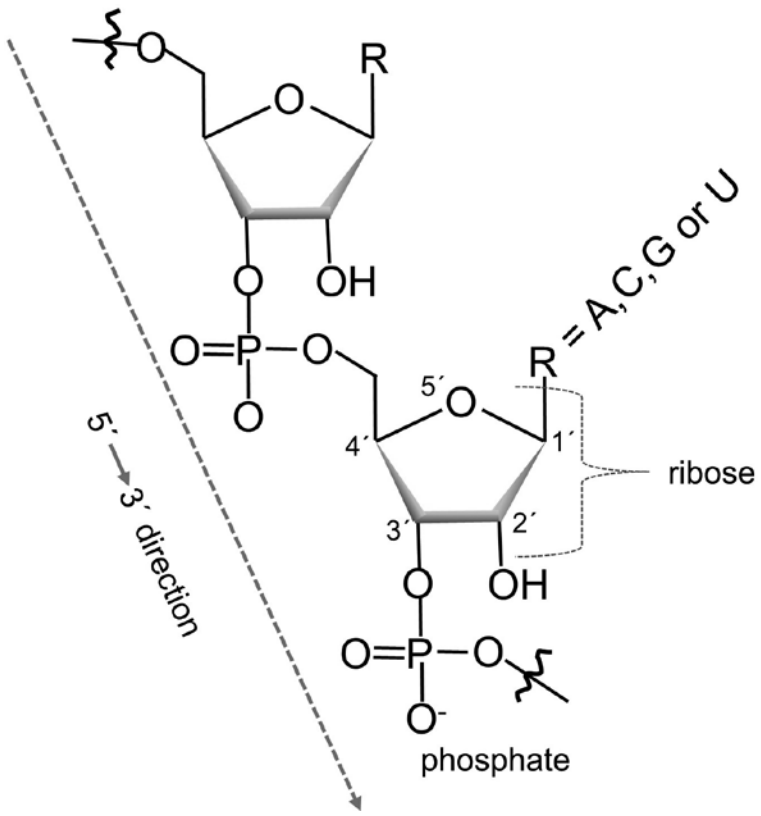


Figure 11. Structure of RNA. The four nucleotides are defined by whether the base A, C, G or U is attached where the R group is shown.

II. Generic Programming Elements

Modern computer languages use some standard constructs to express what is to be done. Often the same logic can be reused, and new values only need to be assigned to the variables. We will discuss the main ones used to implement processing in computer and cellular programming.

Assign Values to Variables

Kirschner and Gerhart noted that information is used by cells to respond to changing circumstances. They wrote, "Two extreme views of information transfer have always existed in biology, the permissive and the instructive. The distinction comes up whenever there is a stimulus and response, or more generally a cause and an effect … Watering a seed provides a stimulus, but it is a permissive input, since no one would assume that the water falling on the seed instructs the seed how to germinate into a plant" (2005, p. 125). We believe their intuition refers to values (provided by the stimulus) and variables (which generate a response upon processing with the assigned value). The cascade of steps to be executed—after, for example, sensing moisture—must already have been prepared and anticipated at the receiving end. The variables patiently wait until activated by informative signals.

Programs and subroutines use variables restricted, as we mentioned, to a relevant data type, to which different values can be assigned every time the program is executed. To illustrate, *price*, *discnt*, *p*, *d*, and *newpri* are variables in this Fortran-like programming code.

```
price = 100                    (5)
discnt = 5
call calc1(price, dis-
cnt)
subroutine calc1(p, d)
newpri = p - d
```

Values have been assigned or are calculated. Here *price* and *p* have the same meaning, and two coding conventions are linked by associating a variable from the calling program to one used within the receiving subroutine *calc1*.

How do variables relate to the discussion on symbols, data types, subsets, and operations above? In computer programs, variable names and their values are constructed from one or more fundamental symbols. The **variable** *price* is defined by combining several symbols from the relevant ASCII alphabet and is treated as a unique entity. The symbol combination *ecirp*, however, has not been assigned a meaning in (5) and is not a valid variable in this program. The **value** 100 assigned to price is also comprised of several ASCII symbols, which taken together have a unique meaning, but assigning price = e34/$![makes no sense, being outside the relevant data type. An operation newpri = TRUE / 45 is also not legitimate, not being a valid operation of that datatype.

Through such precise software conventions, programming errors can be avoided and action to perform expressed unambiguously. However, if the semantic meaning of the variables is not known, the ultimate intention and results might never be fully understood. What if the source code is not available at all but only the executable program? By empirically testing variable values, the hidden Boolean logic can still be discerned in principle by the results, an important observation when reflecting on cells.

How does all this work with cells? To understand cellular logic, one must identify four players: the sending coding system; the receiving coding system; and, for both, what the variables are and what provides their values. What is a variable? It is the biological receptor or sensor able to assume alternative values (including a simple "bound"/"not bound" state), which, once activated with a value, leads to a relevant biological response.

Variables are composed of a single symbol or of elementary symbols combined in a unique manner (in computers and cells). Defining variables is necessary to program intention, and cellular variables are identifiable by humans and cellular decoders.

The *A* site of a ribosome is a **receiver variable** (Figure 2), able to accept as values any of the 64 codons or to be empty. To work properly at the ribosome, not any codon will do. It must only accept the value transferred by a specific **sender variable**, which is associated with the relevant mRNA.

As another example, the location on a template DNA being currently processed by a DNA or RNA polymerase is a **sender variable** whose current value is one of the four nucleotides to be communicated to a polymerase decoder. At the end of the growing chain, part of the polymerase defines a **receiver variable,** which needs to know which nucleotide is to be added (the receiver value) (Figure 3).

The general pattern should now be clear. Special locations on sugars, membranes, or proteins are variables that can accept values (ligands or nothing bound), for example, in the histone code. The enzymes that methylate the appropriate histone residue can have many variables of their own—used to first perform their own internal logic—and then a sender variable is assigned a sending value, the ligand it will transfer. Recall that a chain of sender/receivers can be set up.

The discussion above may have suggested that only a few elementary symbols are used along with a handful of values for variables. Unlike computers, which use only elementary 0/1 "bits" grouped into a relatively modest number of unique ASCII symbols, in cells variables and their values rely on different and more complex alphabets for different codes, using many elementary symbols having distinct geometric and electronic properties.

With computers, hardware design is simpler and more reliable if the variety of elementary symbols (bits) and grouped symbols like ASCII letters are restricted. Many of the cellular codes, however, must support a far more nuanced behavior (recall our comments on fuzzy variables and fuzzy values). A very large number of elementary symbols are used, each having three-dimensional electronic and geometric features (as when portions of amino acids within proteins are combined in TFs). This permits rheostat-like or fuzzy-logic outcomes, which can be fine-tuned dynamically.

To illustrate, not only can different combinations of amino acids define the same kind of TF, but nearby attachments and physical conditions like temperature and salinity can affect the quantitative value that gets interpreted once bound to a CRE. Fine differences in the topology of the same kind of CRE—even those having identical nucleotides—can also lead to quantitative differences upon interacting with a seemingly identical TF. This is important to understand how codes can interact synergistically. They can modify the physical geometry of the compound symbols used by other codes.

Assign a Value to a Constant

Values of variables could change very often during execution of a program, such as the next nucleotide value to be processed by a polymerase. Programs also benefit from using constants, which during a relevant time period should not change. Implicit in cellular logic processing are many constants, such as the temperature, amount of energy provided by an ATP molecule, which ensemble of genes are up- and down-regulated for a cell type, and genomic imprinting (in which certain genes are expressed in a parent-of-origin-specific manner).

Boolean Logic

The ability to use If-THEN-ELSE type logic adds immense value to programming, and to understand cellular logic, one must identity what is the *variable* being tested and what provides its *values*. Between 5% and 10% of protein-coding genes in most organisms encode a TF (values for CREs), and these can have multiple binding domains. Only three kinds of domain are known: cold shock, helix-turn-helix (HTH) type 3, and HTH psq (Walhout et al., 2013, p. 67). Interaction of only portions of a domain with a CRE or other biochemicals define the values (Figure 12).

Example (6) illustrates in programming terms the kinds of logic performed.

```
if (CRE_1 = 'val_1')
{do this}                    (6)
else if (CRE_1 =
'val_2') {do something
else}
else if (CRE_1 =
'val_3') {do the follow-
ing}
else {do nothing, or
continue what you are
doing… whatever makes
sense}
```

Checkpoint if-then logic occurs throughout every step of the cell cycle (Shapiro, 2014) checking for genome damage (Ishikawa et al., 2006), nutritional status (Searle et al., 2011), progress of replication (Segurado and Tercero, 2009), DNA replication (Putnam et al., 2009; Nguyen et al., 2010), DNA damage (Huen and Chen, 2010), chromosome alignment on the spindle pole (Nezi and Musacchio, 2009; Musacchio, 2011), spindle orientation (Caydasi et al., 2010), telomere capping (Ciapponi and Cenci,

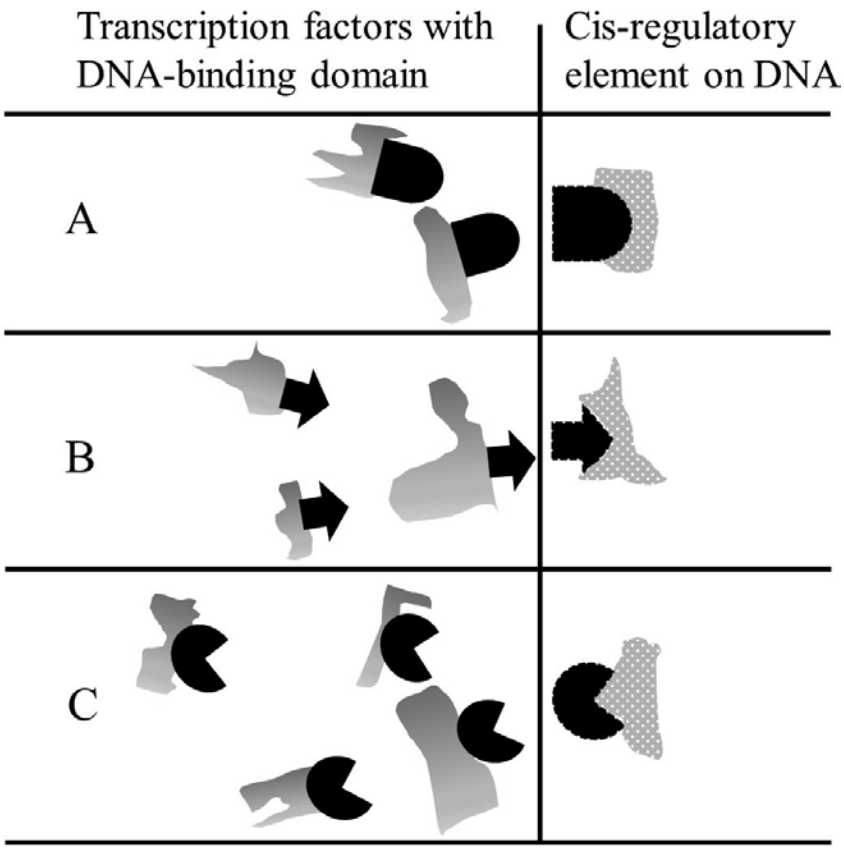


Figure 12. Transcription factors possess DNA-binding domains (solid black), only portions of which provide the values for receiver variables (the appropriate cis-regulatory elements).

2008), cell size (Fang et al., 2006), and whether the cell has accumulated the necessary components needed by the daughter cells (Sabelli et al., 2013).

Errors would lead to serious consequences. Instead of genome repair in response to DNA damage, the if-then logic could lead to programmed cell death (apoptosis) (Tentner et al., 2012; Walsh and Edinger, 2010; Engelberg-Kulka et al., 2009), using some intercell molecules as "death factors" (Holoch and Griffith, 2009) or to a decision to halt the cycle and initiate very sophisticated repairs (Song, 2007).

Iteration

Iteration loops are often used in programming to ensure the correct number of repetitions. An "infinite loop" would consume a computer's—and cell's—resources and must be prevented (Figure 13).

Various repetitive processes occur in cells under the careful regulation of Boolean decisions: many RNA copies are produced from a single gene; many protein copies are made from a single mRNA; many copies of key biochemicals are synthesized, such as amino acids, tRNAs, hormones, ATP, antibodies, etc.; each codon position on mRNAs must be processed; flagella must rotate enough times but not continually; tubulin copies are polymerized to form long microtubules; enough recursive interactions having the right parameters must be run to produce steady-state genetic regulatory circuits; and many copies of each cell type are produced in eukaryotes.

There are many more examples, recognized whenever a cyclic behavior is observed having feedback control. Examination of molecular machines reveals that this is a general principle. Controlling iteration, defining the conditions to use, when to start, and when to terminate, must be implemented simultaneous with the iterating processes. Runaway production would be deadly. Remarkably, this applies not only to the operation of molecular machines but also to the process to create the right number of them also, according to current cellular need. Structuring data into datatypes like arrays and link facilitates the use of iterations in programming.

Control Structures

Programs use techniques to control what is to be done, when, where, how, and how often. In cells, we find many examples. We discussed iteration already. Boolean logic is used with the binding state of cis-regulatory elements (CRE) such as enhancers, silencers, and insulators (Kolovos et al., 2012; Capelson and Corces, 2004) to regulate genes precisely, in a manner unique to each cell lineage (Davidson, 2006). The logic is often very complex. Suites of *cis*-regulatory modules (CRMs) (Figure 5) can regulate multiple genetic loci distributed throughout the genome, establishing network circuits sometimes called "regulons" or "*cis*-regulatory networks" (Dufour et al., 2010).

The combinatorial potential through binding various TFs permits a vast range of regulatory possibilities, able to engage in sophisticated molecular computations (Shapiro and Sternberg, 2005; Davidson and Erwin, 2006). Because the underlying physical interactions are weak, the components can form and dissociate rapidly to permit quick responses to signals received. Complex computations using weak interactions to form novel circuits is also typical of

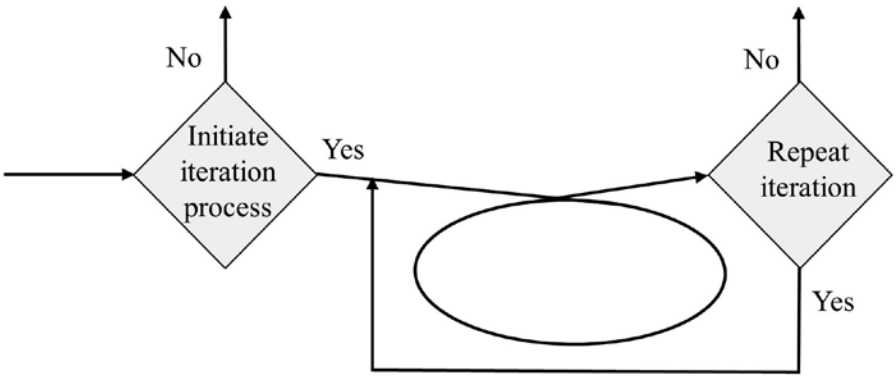


Figure 13. Iteration loops are common in computer and cellular programming. Conditions are tested to determine when to initiate an iterative process and when to repeat or terminate it.

how neurons are wired (London and Hausser, 2005; Sidiropoulou et al., 2006; Markram et al., 2015).

Computer programmers can use "**GoTo**" type commands. Special signals are ubiquitous in cells, which specify where molecular machines and components are to act, i.e., which organelle, subcompartment, or location on a membrane. Causing instructions that are stored elsewhere to be executed goes by names such as functions, methods, procedures, and subroutines in computer programming. In cells, there are many examples, such as activating hox genes to regulate expression of many genes as a modular ensemble and activating key TFs to generate genetic regulatory genetic circuits (Davidson, 2006). Remote processing is often encapsulated in various subcompartments and organelles. We recall that DNA is also present in plasmids, mitochondria, and chloroplasts, not just chromosomes. These decisions also require the use of variables.

Another technique used by computer languages is the idea of "sleep" or "**wait**" for a fixed or variable time period. We find many examples in cells, such as feedback inhibition in enzymatic networks, gene deactivation, and placing the cell cycle on hold.

Other Non-Prescriptive Processing

Most of what happens in computers results from explicit instructions, but our analysis of coded information systems clarifies that additional physical constraints are also always incorporated to ensure the intended outcomes. There are design trade-offs, whether to guide intention as coded messages or in a hard-wired physical manner. A computer example is when printed paper falls into a tray with sides that hold them in place. A considerable amount of cellular success is based on pure physical-chemical factors that have been carefully organized, a topic we discuss in Part 2.

Read and Write

Computer programs read, write, and delete to long-term and short-term memory devices. The codes found in cells must be able to read and write data values. Setting epigenetic tags are examples of medium and long-term write operations, which serve to communicate intended outcomes later. DNA is usually thought of as a fairly permanent source to read data from, but DNA can be added to a genome via CRISPR (Zetsche et al., 2015; Ran et al., 2015; Gen News Highlights, 2015), reverse transcription (e.g., telomerase reverse transcriptase that maintains the telomeres of eukaryotic chromosomes), transfer and acquisition of new genes via integrons coding cassettes (Hall and Collis, 1995), and different lateral gene-transfer mechanisms, including transfer of plasmids, in prokaryotes. Inteins are another mechanism. These are self-splicing portions of proteins with homing endonuclease ability that snip parts of DNA so that a copy of the coding sequence of the intein can be inserted there (Gogarten et al., 2002).

DNA can also be modified in other ways. DNA segments such as transposons can be transferred to other sites on the genome, and "shufflons" can invert sections of DNA, for example, to replace part of a coding strand with its complementary strand to create modified proteins (Tam et al., 2005; Komano, 1999).

Multiprocessing and Threading

Modern computer hardware and software designs can parallelize computations, permitting multiple tasks to be carried out simultaneously. This is common in cells, such as in the parallel production of ATP from many mitochondria; translation of several identical mRNAs in parallel (several ribosomes can also translate the same mRNA simultaneously), transcription of multiple copies of the same gene, the existence of many cells of the same kind, and the presence of multiple copies of the same subcompartments and organelles.

Reuse of Modules

In good software design, the same general-purpose modules, methods, and procedures are often reused. A common approach is to separate identical portions of coding into smaller modules that can be invoked from within overarching modules. This modularity is found

also in cells. As Kirschner and Gerhart pointed out (2005, p. 137), "The same pathways are used over and over again within the same organism for different purposes. Thus, they must be modified slightly to interact with a variety of processes and to work in different environments and cell types." They describe the interactions as "weak linkages," which we recognize as simply variables or parameters used to link subprocesses in different manners.

Interchangeable Libraries

In addition to invoking subroutines, sections of computer code such as classes are often imported from a library. Similarly, prokaryotes in particular exchange genetic material through horizontal (lateral) gene transfer (Thomas and Nielsen, 2005; Ochman et al., 2000; Koonin, et al., 2001), whereby genes, plasmids, and so called "islands" encoding specialized adaptive functions are exchanged (Dobrindt et al., 2004). This permits a huge amount of coding to be distributed in the environment and put to use rapidly when the need arises, facilitating adaptability. This is a form of open systems design. Genetic material can also be transferred into eukaryotes through vectors such as viruses.

III. File formatting

Shapiro and Sternberg (2005) drew attention to the parallels between computer file formatting and data storage in cells:

> The explicit parallel with electronic data systems indicates that the genomic storage medium has to be marked, or formatted, with generic signals so that operational hardware can locate and process the stored information. (Shapiro and Sternberg, 2005)

Data storage can be organized physically in computer and cell technologies using principles such as sectors, disk partitioning, and data segments, discussed in Part 2. On top of this infrastructure, software programs organize different data using file formatting. A program that interacts with specially structured file data must be able to access it correctly, even though the location of the content could be scattered all over the physical medium. DNA, RNA, and proteins are used as read/write/delete storage devices and need to be properly formatted so the corresponding "reader" will work.

The metadata contained in a computer file header can be stored at the start, end, or other areas of the file. Likewise, in DNA, RNA, and proteins formatting instructions need not be found in only one location. Given the existence of multiple codes, DNA "files" are formatted for use in different manners, depending on the program being used. The various ways DNA are packed, such as by nucleosomes, determine which genes can be processed. Preparing portions of DNA for processing by DNA polymerase (to identify the starting and end points, open and unwind the strands, remove bound histones, etc.) is very different from the formatting details—which occur in three dimensions—for RNA polymerase. The programs that perform DNA error corrections also require their own formatting rules.

Epigenetic tags are often used to identify what data to process and how. Adding and removing these ligands from DNA, RNA, and proteins is an example of preparing files for processing and must be carefully regulated. Histone modifications define which portions of DNA can be processed. Over a hundred posttranscription modifications have been identified in all three major RNA species (tRNA, mRNA and rRNA), as well as in other families of RNA such as snRNA (Cantara et al., 2011). Examples of formatting specifications in DNA include the use of methylation and demethylation (Bird, 2002; Paszkowski and Whitham, 2001), binding of TFs (Cheng et al., 2012; Davidson, 2006), and rules to identify exons (Harrow et al., 2009).

Individual eukaryote mRNAs are formatted as individual files with beginning and ending metadata in the form of 5' capping and 3' polyadenylation, attached miRNAs, and so on. This is necessary to ensure the ribosome program will work properly. Different sets of formatting rules are necessary for different programs such as separation of introns and exons by the spliceosome or to degrade RNA.

Formatting on proteins is common. Posttranslation modifications (PTM) include methylation, phosphorylation, acetylation, ubiquitylation, glycosylation, and sumoylation (Strahl and Allis, 2000; Jenuwein and Allis, 2001). Structural three-dimensional recognition features, generated with alpha coils, beta sheets, disulfide bonds, hydrophobic patches, and other features also ensure correct formatting of proteins. In cells, all this is precisely regulated, often down to the atomic level. Reversible phosphorylation, the most widespread PTM, occurs on the correct atom of a serine, threonine, or tyrosine residue to form phosphomonoesters or on histidine, arginine, and lysine residues to form phosphoramidates (Cieśla et al., 2011), all according to the particular code involved. Recalling the existence of signal cascades and enzymatic networks, proteins are also carriers of data values that get processed by other sensors (variables to be assigned values). DNA and RNA are not the only information carriers in cells. A modification on a TF can become a data setting to be used by the receiving portion of a second TF. For these reasons we see that proteins can be formatted and classified into different "file types."

Copies of tagged proteins, RNA, and DNA (like nucleosomes) can be inherited by somatic daughter cells, and sometimes the tag is removed from the daughter cell, generating an empty or partially empty "file" that can be written to. In the same way that a program like Excel cannot process a jpg file, each

of the cellular information "readers" process only the data specially formatted for it.

Compressed Archival

DNA is compressed and protected for future use by winding sections of ~ 147 base pairs around a core of 8 positively charged histone proteins into nucleosomes, and then further compacting the nucleosomes into higher order chromatin structures complexed with protein and RNA (Jenuwein and Allis, 2001). The portions of DNA that need to be expressed must be unpacked and reformatted properly. The cellular goal is to save physical space and protect the medium from degradation. Bacteria also quickly lose DNA not immediately needed (and can regain genes via lateral gene transfer), which saves precious raw materials and energy. Computer analogies of the principle include programs like zip and the export of tables by a database management into a single export file, all or parts of which can be retrieved and properly structured for use later. However, computers use algorithms that recode the original content using fewer bits, a principle not known in cells. Inspired by cellular compression, which transforms essentially linear DNA into three-dimensional storage, engineers might consider designing mass storage devices to also store data that cannot be used immediately as is but, like packed DNA, could be reopened when needed.

Summary and Discussion

Recognizing cells as information processing devices is the proper way to understand their holistic intent and design. In fact, Gatlin (1972, p. I) defined life as an "information processing system," and Britten (Britten, 2003, p. 82) pointed out, "We cannot start with DNA and grow a cell because there must be an adequate initial state of a cell with a vast multitude of details under control." We mentioned above that cellular information is partially distributed hierarchically and recognize that there are many carriers in the lower, embedded levels. An organ consists of many cells, each of which contains many mitochondria, and so on. In large populations of prokaryotes, the logic processing is distribution over many interacting species to form a viable ecology, whereas in complex eukaryotes considerably more is concentrated within the individual organism. In virtually all biochemical processes, one sees strong regulation unless the process is malfunctioning, as in cancerous growth or viral infection. In other words, there are always sophisticated rules for when to begin and countermeasures that prevent runaway processing.

Regulation is best designed and interpreted using purely formal rules, a key feature of software engineering. If, for example, a metabolic chain requires feedback control to a preceding enzymatic reaction, this can be analyzed and expressed symbolically, along with the mathematical specifications and control rules. To instantiate the requirements, a physically viable solution then needs to be implemented. No rational engineer or programmer would think of developing programs by letting rules and their implementation pop into existence randomly without any conceptual guidance.

We saw that conceptual software elements such as iteration and control structures are developed on top of data types—each with their unique properties—organized into variables, arrays, and linked lists and all this using well-defined file formatting to facilitate processing by molecular machines. Many independent codes found in cells make use of these principles. It is hard to overstate how important variables are in cellular processes to permit regulation and maximum adaptability. The location, timing, and amount of transcription by RNA polymerase is defined by CREs (promoters, enhancers, silencers, insulators; Kolovos et al., 2012) and termination by terminator sequences (Ishihama, 2000).

It would require many volumes to describe in detail the formal control structures used by other cellular activities, such as homologous chromosome crossing-over, VDJ recombination in the immune system, nonhomologous end-joining (NHEJ) of broken DNA ends, DNA transposons (self-insertion, excision), telomerase extension, chromosome segregation, DNA compaction, binding sites affecting DNA spatial organization into transcription factories in the nucleus, signals for error correction and damage repair, and the multitude of other cellular processes.

There is considerable evidence that damage through random changes is actively hindered in cells, such as a bias for many retrotransposons to insert upstream of transcription initiation sites (Shapiro and Sternberg, 2005), which prevents damage to coding sequences and enhances the potential for a constructive regulatory change. Very often the regulatory logic makes sense to humans skilled on symbol logic, but the details are different across taxa and did not originate from a common ancestor. An example is the signal used in E. coli to repress catabolism (the CRP palindromic binding site for the CRP-cAMP complex), which is unrelated to that found in Bacillus subtilis (CRE element recognized by protein CcpA) (Miwa et al., 2000).

Coded Systems Can Interact

Although the various codes operate independently in cells, they can collaborate to ensure a fine-tuned outcome. We mentioned epigenetic codes, which modify gene expressions, and another code based on TFs bound to CREs, which also regulate gene expression. But in addition, a different code based on adding and removing ligands—especially phosphate groups—modify the TFs themselves (Shapiro 2006). Furthermore, TF half-lives are also regulated

by the NEnd code. Gene expression is further affected by other codes which use various classes of RNAs (siRNA, snoRNA, miRNA, etc.) that modify chromatin accessibility, transcription initiation, transcription elongation, RNA processing, RNA stability, and mRNA translation (Mattick and Makunin, 2006; Taft et al., 2010; Storz and Wassarman, 2005).

By integrating multiple codes, cells become highly responsive to what is going on throughout the entire cell and their external environment. The design requirements would be overwhelming for humans. The same stretch of DNA can be used as variables for some codes (e.g., CREs, methylation binding locations, and after transcription to locate regions on mRNA for miRNA binding and to specify intron/exon boundary locations) while simultaneously providing the data values for other codes (e.g., as codons after transcription and as a template for new DNA copies).

These requirements demand formal specifications to satisfy all requirements and to define what is to be done by each code. Synonymous coding from the point of view of the genetic code must identify protein sequences while simultaneously controlling translation rates within regions of mRNA. The DNA-to-RNA conversion code during transcription also needs to control stalling of mRNA precursors for spliceosomes for purposes of siRNA accumulation as part of a host's defenses to damaging transposons (Dumesic et al., 2013).

Collaboration between coding systems is sometimes linked directly. The histone modifications, which involve over 100 protein readers, writers, and erasers (Carey, 2012, p. 72, 224), sometimes develop protein complexes that include the enzymes that methylate CpG motifs on DNA (DNMT3A and DNMT3B) in the same region the histone is located (Carey, 2012, pp. 73, 89–90). This is another example of instantiation using adaptor molecules.

The reverse is also true. The DNA methylation code can affect the histone code in a synergistic manner. Methylation attracts more repressive histone modifying enzymes (Carey, 2012, pp. 224–226). Similarly, long ncRNAs locate near imprinted genes (which identify whether coming from the mother or father), and these can recruit epigenetic enzymes such as G9a or EZH2, which put a methyl tag on lysines K9 and K27of histone H3 (a second code) to enhance the imprinting (Ikegami et al., 2009). To complicate the picture, long ncRNAs can increase or decrease expression of target genes for reasons not understood.

The miRNA code also interacts with enzymes involved in epigenetic codes by regulating their effective concentration (Carey, 2012, pp. 231–232).

Stem cells express a very different set of proteins than differentiated lineages. Not only are different genes deactivated by blocking TFs bound in the cis-element region, but also a different set of miRNAs are switched on (a second code) to help identify and degrade the mRNAs no longer needed by that class of cells (Pauli et al., 2011).

Chemotaxis (ability to swim toward nutrients and away from noxious stimuli) uses two codes in E. coli to respond to more than fifty substances. In the first one, there are four kinds of receptors on the membrane that respond to the environment by phosphorylating the communication protein CheY, which can modify the direction of rotation of the flagellar motor through binding at certain locations. A second code affects the four kinds of receptors themselves by adding and removing methyl groups to any of eight different sites per receptor. The receptors are grouped into triplets on the membrane, so the number of possible methylation states is astronomically large. The net outcome of these two coded processes is to permit the bacteria to "in effect perform calculus" (Bray, 2009, p. 94). It is not the absolute concentration of external stimulant that determines the decision to change direction of movement, but rather a large change in the relative concentration (Bray, 2009, pp. 89–97).

In Part 2, we flesh out our understanding of cells as holistic entries whose hardware components must also be taken into account in addition to the interacting codes. It is wrong to think DNA provides detailed instructions on how to assemble an organism. Oyama (2002) pointed out that "a gene initiates a sequence of events only if one chooses to begin analysis at that point: it occupies no privileged energetic position outside the flux of physical interactions" (p. 40) and that "gene transcription and translation in no way represent instructions for building a functioning body" (p. 69). She correctly mentioned that the interactions needed to define organisms are inherited as already functioning cells and in a similar environmental context as the parent (pp. 17–18, 26, 43–49, 77).

Dynamic Nature of Cellular Control

The location of data in computer memory is rearranged in controlled manners and address pointers are used to identify the location of data. For cells this is also true, but the process is more sophisticated. A TF can search for a CRE in three-dimensional space and is robust to physical degradation of its target through mutations. Unlike a computer pointer to a single address, in cells *n* identical TFs or other signals can point to multiple locations to activate an ensemble of process-related genes. In computers, a memory address is usually referenced directly, whereas in cells often a linked chain of pointers referencing other pointers lead to the sites to be activated, which permit refinements, including fuzzy logic (Kosko and Isaka, 1993; http://zadeh.cs.berkeley.edu) to be integrated at every step.

Analog Computers

We have not mentioned principles from the less-known analog computers in this

introduction to logic processing in cells. We only wish to point out here that the wide diversity in sensors responding to signals can produce a rheostat-like response (i.e., a continuum of response). Software designed for digital computers would process this kind of logic by defining ranges of values for these variables and program the appropriate behavior for each range. This relates to our suggestion above that computer scientists consider using fuzzy variable and fuzzy values, being a principle cells use.

Neo-Darwinism Fails to Explain the Origin of Logic Processing

In *The Plausibility of Life*, we read, "The architecture of cells is achieved without an architect. No central regulation is discernible. Cells are in fact capable of many structures; many are chameleons that change their structure in response to circumstances" (Kirchner and Gerhart, 2005, p. 148). It is correct that there is no set of instructions on DNA that specify the detailed order in which events are to unfold, but this does not deny an architect; in fact, it indicates a creator who designed for adaptability to changing circumstances (Truman, 2015). As mentioned above, a virtually unlimited variety of responses can be executed by using enough variables and their values. Adaptability is found everywhere in biology, not only within cells. Gilbert (2003) provides several examples of dramatic polyphenism, or open systems adaptability, such as sex determination of blue-headed wrasse larva depending upon the presence of other males or females nearby; diet in caterpillars, which enables them to change their morphology to camouflage themselves according to season when born; and predator-secreted chemicals.

Cellular process must be initiated and stopped. Runaway execution would rarely if ever be acceptable, but why should the termination rules develop in advance of these thousands of formal logic-guided processes? Which evolved first, the process or the means to turn it off? Natural processes cannot look ahead to plan complex solutions to make cells and entire organisms adaptable. Gene regulatory networks, signal cascades, metabolic networks (Figure 14), and the operation of molecular machines are regulated at many levels using programming constructs recognizable by human designers.

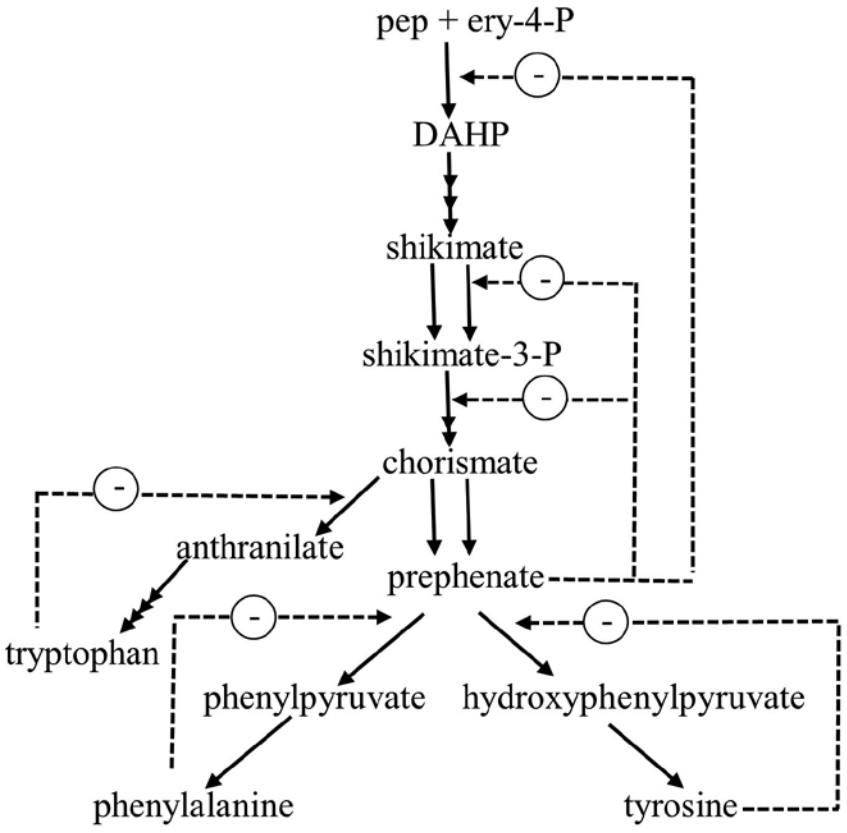


Figure 14. Enzyme chain including feedback in aromatic amino acid synthesis (Fell, 1997, p. 209).

There is no analogy in inanimate matter of codes being used to express an intended result to ensure continued system integrity. This will become clear after examining in the next paper how extraordinarily complex the molecular machines are which are needed to implement the code specifications. In Figure 15, we clarify the principle, which is not found anywhere in inanimate nature.

The intuition is that a system with complex internal components will be repetitively confronted with a decision that can be freely made, independent of chemical or physical compulsion. For each iteration a particular choice between alternative paths is correct to facilitate the survival of the system (plus the decision-making apparatus), based on current circumstances.

The cell is full of this decision principle, such as where to initiate and stop transcription, which amino acid to add next to a growing protein, and where a restriction enzyme should cut.

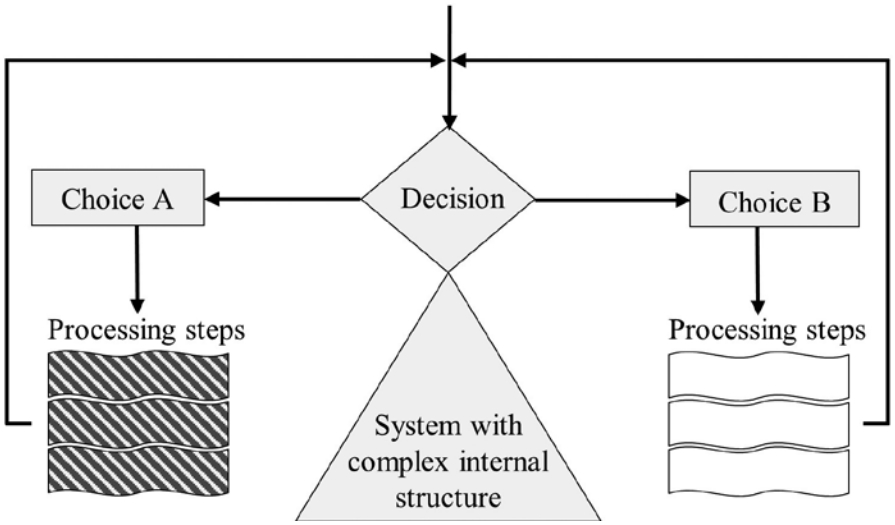


Figure 15. In inanimate nature we find no examples of systems with complex internal structure repetitively facing a contingent decision and then making the correct choice for each iteration based on which outcome supports survival of the system during that iteration.

We do not find any examples in inanimate nature, even though this is but a minimalist requirement. We are not demanding this occurs reliably a huge number of times (like millions of correct peptide bonds during a cell's lifetime) or that it be able to manufacture all its key components, or that the entire apparatus be reliably replicated for many generations. We ask only for examples showing the basic concept is found in some elementary way in inanimate nature. Otherwise, no evolutionary theory is justified in simply assuming grotesquely more complex cells arose in the absence of intelligent guidance. This is essentially asking about the origin of information, like the sequence of codons specifying the correct protein chains.

We know from computer technologies how important proofreading and error correction (parity bit rules, etc.) are during data storage and transmission. In cells, this is far more important, given the many examples of iteration and millions of decisions per second involved. We will consider just one code, the genetic code, to illustrate the need for extreme reliability. If the multiple copies of mRNA and their translation products were error prone, this would lead to error catastrophe during the cell's lifetime. Each new batch of flawed proteins and RNA would lead to ever more defective transcription factors, RNA polymerases, ribosomes, spliceosomes, error-correcting enzyme complexes and posttranslation machines, thereby producing ever more defective proteins and RNA the next time around. The same, of course, is also true about all the components inherited by daughter cells, in particular flawed DNA copies.

Is this a serious problem? The probability that an amino acid will be translated correctly depends on many factors, but suppose that in the distant evolutionary past, before elaborate error-correcting molecular machines existed, natural processes had somehow miraculously reached a state where each of the twenty amino acids was translated correctly with an average probability of 0.80 and that proteins back then were on average only 200 residues long. The chance of obtaining a correctly translated protein would be $(0.8)^{200} = 4x10^{-20}$.

One recent study of 40 proteins examined in HeLa cells concluded that the lowest number of copies per cell at a given time was for the oncogene FOS (6000 copies), and the most abundant was vimentin (20 million copies) (Zeiler et al., 2012). An ancient primitive organism would not have so many copies. We would not expect to get even one correctly translated protein but a sea of hopelessly flawed, misfolded, and destructively interacting ones (for a more exact analysis see Part 2). Even if the cell could somehow recognize and degrade mistranslated ones (somehow using molecular machines that themselves are hopelessly corrupted), the energy cost to produce enough attempts to generate thousands of necessary good copies would be prohibitive.

What is the reality in all cells studied? Success rates on the order of "only" 0.8 per monomer copied? Many processes recognize and correct errors, such as when DNA is replicated or tRNAs are charged. In exonuclease proofreading during DNA replication, a mismatched duplex is identified and the most recently incorporated nucleotides removed and replaced, eliminating about 99.9% of accidental misincorporations from the nascent strand (Kunkel Bebenek, 2000; Ibarra et al., 2009). A second mechanism, postreplication mismatch repair, then corrects about 99% of those misincorporations that escape exonuclease proofreading (Modrich and Lahue, 1996; Kunkel and Erie, 2005). There is also a molecular machine to repair double-strand (DS) breaks (Brissett and Doherty, 2009).

The other codes must also be highly accurate. TFs could bind to a multitude of wrong locations on DNA; epigenetic

tags on histones or DNA must be carefully controlled; flawed signal sequences would cause proteins to be secreted improperly; etc. Furthermore, the coded variables and variable values must be replicated accurately over many generations, not just the organism's lifetime. A distinct combination of millions of methyl tags on DNA cytosines is unique to each cell type and needs to be replicated in daughter cells (Carey. 2012, p. 60) as shown in Figure 16.

We do not know how accurately the methylation pattern per CpG must be replicated for the daughter lineages to still work, but suppose it would be enough if "merely" 1/10,000 need to be correct (i.e., 99.99% error rate would not matter, the resulting pattern would still work). Per replication and one million CpGs, a successful outcome would only occur $4x10^{-44}$ of the time (i.e., $0.9999^{1000000}$), even given such generous constraints. In other words, getting an acceptable copy will not occur even if only 1/10,000 tag positions need be correct on average. We conclude that evolving this new function cannot start crude and be refined by random mutations, since natural selection would have nothing functional or consistent to work on.

Many researchers, especially those of a neo-Darwinian persuasion, continue to downplay the evidence for deliberate planning found in cells, preferring to hold on to the myth that most the genome is junk instead of facing the reality of multiple codes and an overarching systems design. The origin of complex features is assumed to result from random mutations followed by natural selection without recognizing or addressing the origin of formal logic processing (Dawkins, 1996). Absent informational guidance, the only alternative is to believe in a series of naturalist miracles, such as an initial functional genetic apparatus followed by many more miracles including a regulated energy source (ATP molecules) and requirements such as being able to distribute chromosomes and other components to daughter cells.

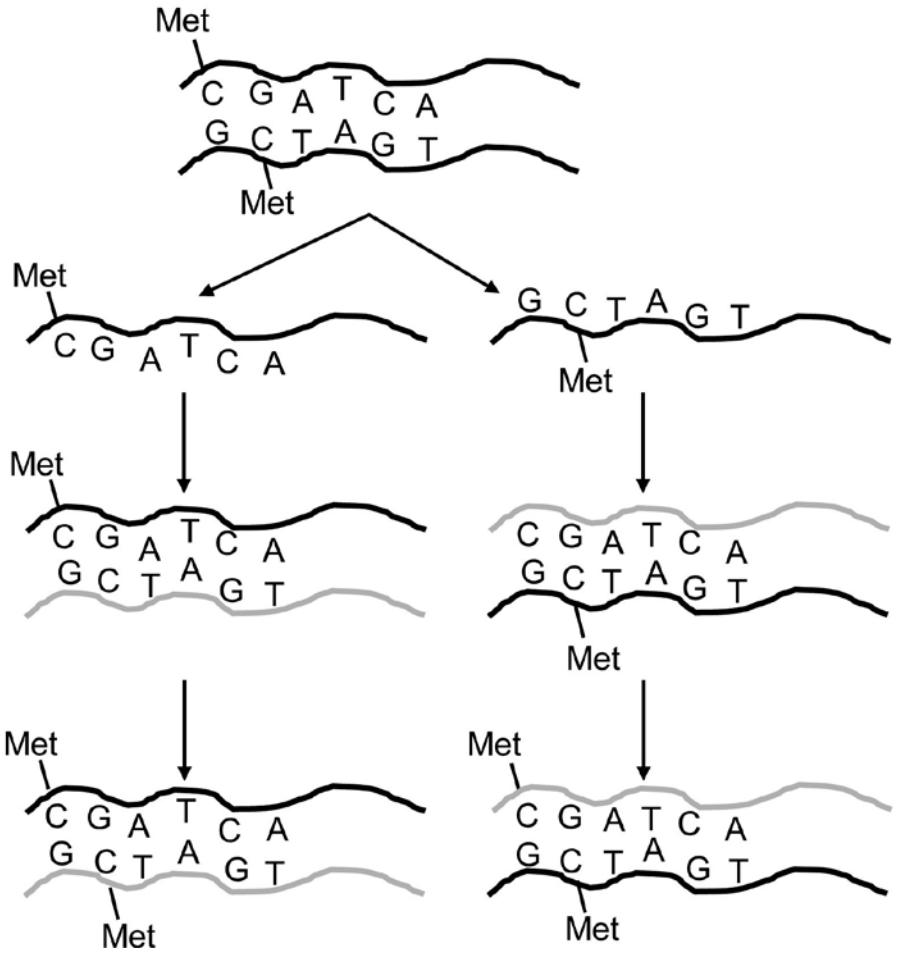


Figure 16. DNA methylation patterns need to be replicated in daughter cells during somatic cell division. After each DNA strand is separated and the second strand copied, the DNMT1 enzyme searches for CpG motifs and transfers a methyl (Met) group to the new strand where needed. This results in two new copies carrying the original methylation pattern.

Is this also a probabilistic nightmare? There are $2^{(46 \times 2)} = 5x10^{27}$ ways to distribute human chromosomes during mitotic cell division (Page and Hieter, 1999), of which only one is correct. There is a better chance to guess two people correctly in a row out of everyone who ever lived. And these odds need to be overcome by every surviving cell every generation, so once again error cascade is the natural consequence until the process is close to flawless. Natural selection is only relevant once the system has attained miracle-level perfection.

In general, whenever we come across the terms "convergent evolution," "genetic piracy," or "co-optatation," we will discover a failure of neo-Darwinian theory and in all likelihood further evidence that logic processing elements are being deliberately reused in unrelated organisms. For many years the very small

amount of data available was misused (and continues to be) to claim that a gene expressed as part of the same or similar processes reveals common ancestry. In the words of Striedter (Striedter, 2003, p. 287), "Unfortunately, we now know that most genes are expressed in several different locations and that many homologies based on the expression patterns of single genes have turned out to be controversial, to say the least."

Until one accepts that cells are designed logic processors, much data will continue to be misunderstood. The same transcription factor or the same cis-factor pattern could be reused for biologically unrelated purposes across the biosphere. In programming, we also find software elements such as "for (int *i* = 0; i < *myList.size*(); *i*++)" in many programs, but this does not imply the programs are related in any manner. The *i* and *myList* could represent totally different things.

In discussing the Pax-6 gene found in vertebrates, Drosophila, squid, and even flatworms, Willmer provides an example: "Although this could imply a common starting point for all eyes, it is more likely an example of the universality of positional and pattern-forming determination systems in animals. Note also that while Pax-6 in vertebrates is homologous to the Drosophila gene eyeless, other genes related to eye formation in vertebrates match bizarrely with genes involved in appendage formation and with muscle formation in fruit flies; and that Pax-6 also regulates the unrelated phenomenon of nasal placode formation in vertebrates" (Willmer, 2003, p. 38).

Premature evolutionary speculation, treated and repeated for decades as proven scientific fact, is being increasingly corrected. Discussing the claim that the gene *engrailed*, which is expressed in both *Drosophila* and chordate metameres, proves that segmentation of body parts goes back 500 million years ago to a common ancestor, Willmer explained what more data now actually reveals: "This now seems an overinterpretation. Although homeobox proteins function as transcription factors for other genes, the genes they regulate are often quite unrelated to segmentation. Furthermore, this same Hox sequence appears in a far greater range of animals, including unsegmented nematodes and echinoderms" (Willmer, 2003, p. 39). After providing other examples, Willner then arrives at the correct intuition: "The similarity of genes … may lie in processes rather than in real homology" (p. 40).

Scientific Guidance through the Design Presupposition

The NIH Roadmap Epigenomics Consortium is collecting a huge database with DNA accessibility, RNA expression, histone modification, and DNA methylation patterns for 111 human reference epigenomes (Kundaje et al., 2015). One goal is to identify regulatory modules that arise during cell lineage specification and differentiation. This is representative of the general direction modern cellular research is beginning to take, where it has become indispensable to apply principles from symbolic logic processing to understand in detail the design of cells. Speculative neo-Darwinism is at best post-facto storytelling; it provides no insights into the big, interesting biological questions.

The view that cells were deliberately designed to be robust and adaptable for long-term viability and interactivity, along with the insights of logic processing principles from computer programming, stimulates many fruitful ideas to guide future ideas that do not arise from the evolutionary worldview. Freed from the shackles of possible biological functions being constrained to what a primitive common ancestor initially provided and a limitation on mutational accidents to generate nontrivial novelty, we suggest how our paradigm provides value to guide future research.

1. Cells will be found to be more adaptable than suspected to situations not encountered before, and when the mechanisms are researched, we will find the adaptive logic has coding aspects, meaning the variables were already there and able to process additional values. Asking how one would formally design an optimized outcome, independent of any misguided prejudice from common ancestry constraints, should help identify new cellular control processes. (Post-facto claims for unexpected "convergence" is scientifically worthless and contradicts neo-Darwinian expectations.)
2. Many more forms of complex regulation remain to be discovered than suspected. No iterative process will be found that lacks a formal set of rules on how to initiate and terminate (unless malfunctioning). Whenever it would make sense for the concentration and distribution of biomolecules to vary, we predict evidence will be found this has been implemented in a context-appropriate fashion.
3. Given our conviction that cells were designed to function as holistic and integrate entities, we predict ever more discoveries of interconnectivity between codes so that inputs throughout the entire cell and ecosystem can be taken into account to regulate processes optimally. We expect much will become clear only as the optimization trade-offs are understood and that quantitative analysis will reveal there could not have been nearly enough evolutionary trial-and-error attempts to explore and fine-tune these optimized trade-offs.
4. More quality control checks will be discovered at key processing points. Researchers should search for error checks/correction during transcription to RNA and other key interfaces. Considering the value to cells of recycling valuable raw materials of every kind, we anticipate novel

discoveries designed to ensure this. Conversely, if substances (like cyclic RNAs) are found to be long-lived, we suggest the Creator had a biological reason.

5. We expect that when important alternative pathways are available, the overall optimal one under those circumstances is selected unless clearly malfunctioning. As an example, whether to attempt error correction or initiate apoptosis is a significant decision for cells based on complex cost/benefit/risk tradeoffs. We expect a careful quantitative application of decision theory principles—including Bayesian statistics—to reveal that the outcome selected is overall rational.
6. For every difficult step creating a critical potential processing bottleneck, mechanisms will be found that resolve these, in the same way that we expect that an enzyme will be found to catalyze all key biochemical reactions impacting the survival of a cell. We also anticipate that variants of current enzymes and processes can easily be generated when it makes sense. This is based on our view that general-purpose solutions were often designed, which like good open systems design, are adaptable. Optimized adaptability has nothing to do with the naturalist assumptions going under the label evolution.
7. Since we believe organisms were created optimally (with the goal of filling the earth's ecosystems) but have accumulated errors over time, we will discover residual evidence for functioning solutions in the past, at the cellular or higher level, which do not work as well as before, especially for organisms that have undergone population-size bottlenecks. Applying design reasoning to describe how ideal solutions would work will help us understand how things might have worked before.
8. We will discover multidimensional forms of data storage and retrieval not known for computers. These will be sophisticated beyond anything a naturalist would dare predict. We anticipate the existence of extraordinary code-based methods to store, retrieve, index, network, and consolidate in fuzzy logic and other mathematical forms all kinds of multimedia data (smell, vision, taste, sound, tactile memories, reasoning chains, numbers, facts, etc.) in ensembles of brain cells. We dare predict human minds will be found to be able to interact with these codes in read/write fashion to actively guide queries in a parallel processing fashion.

Acknowledgments

We wish to thank Dr. Peer Terborg and Dr. Niko Winkler for many fruitful discussions. We thank Dr. Michael Heisig for considerable help obtaining much of the literature cited here, Wort und Wissen in Germany for many conferences and opportunities to discuss these topics, Creation Ministries International for years of fruitful collaboration in this area, and Jonathan Bartlett for a detailed review of an earlier draft.

References

André, A., H. Kaltner, J.C. Manning, P.V. Murphy, and H.J. Gabius. 2015. Lectins: getting familiar with translators of the sugar code. *Molecules* 20:1788–1823.

Aricescu, A.R., and E.Y. Jones. 2007. Immunoglobulin superfamily cell adhesion molecules: zippers and signals. *Current Opinion in Cell Biology* 19:543–550.

Ashcroft, S.J. 1997. Intracellular second messengers. *Advances in Experimental Medicine and Biology*. 426:73–80.

Audit, B., and C.A. Ouzonis. 2003. From genes to genomes: universal scale-invariant properties of microbial chromosome organization. *Journal of Molecular Biology* 332:617–633.

Barbieri, M. 2003. *The Organic Codes. An Introduction to Semantic Biology*. Cambridge University Press, Cambridge, UK.

Barbieri, M. (ed.). 2009. *The Codes of Life: The Rules of Macroevolution*. Springer. (See especially chapter 8 by Mario Gimona.)

Barash, Y., et al. 2010. Deciphering the splicing code. *Nature* 465:53–59.

Bass, B.L. 2002. RNA editing by adenosine deaminases that act on RNA. *Annual Review Biochemistry* 71:817–846.

Bernardi, G. 1990. Le génomes des vertébrés: organization, fonction et evolution. *Biofutur* 94:43–46.

Bird, A. 2002. DNA methylation patterns and epigenetic memory. *Genes and Development* 16:6–21.

Bogdan, C. 2001. Nitric oxide and the regulation of gene expression. *Trends in Cell Biology*. 11:66–75.

Bonifacino, J.S., and B.S. Glick. 2004. The mechanisms of vesicle budding and fusion. *Cell* 116:153–166.

Bray D., and T. Duke. 2004. Conformational spread: the propagation of allosteric states in large multiprotein complexes. *Annual Review of Biophysics and Biomolecular Structure* 33:53–73.

Bray, D. 2009. *Wetware: A Computer in Every Living Cell*. Yale University Press, New Haven, CT.

Briscoe, J., A. Pierani, T.M. Jessell, and J. Ericson. 2000. A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. *Cell* 101:435–445.

Brissett, N.C., and A.J. Doherty. 2009. Repairing DNA double-strand breaks by the prokaryotic non-homologous end-joining pathway. *Biochemical Society Transactions* 37:539–545.

Britten, R.J. 2003. Only details determine. In Müller, G.B., and S.A. Newmann (editors), *Origination of Organismal Form: Beyond the Gene in Development and Evolutionary Biology*, pp. 75–86. MIT Press, Cambridge, MA.

Cantara, W.A., P.F. Crain, J. Rozenski, J.A.

McCloskey, K.A. Harris, X. Zhang, F.A.P. Vendeix, D. Fabris, and P.F. Agris. 2011. The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Research* 39(suppl 1): D195-D201.

Capelson, M., and V.G. Corces. 2004. Boundary elements and nuclear organization. *Biology of the Cell* 96(8):617–29.

Carey, N. 2012. *The Epigenetics Revolution: How Modern Biology Is Rewriting Our Understanding of Genetics, Disease and Inheritance*. Icon Books Ltd., London, UK.

Caydasi, A.K., B. Ibrahim, and G. Pereira. 2010. Monitoring spindle orientation: spindle position checkpoint in charge. *Cell Division* 5:28.

Cech, T.R. 2002. Ribozymes, the first 20 years. *Biochemical Society Transactions* 30(6):1162–1166.

Cessac, B., H. Paugam-Moisy, and T. Viéville. 2010. Overview of facts and issues about neural coding by spikes. *Journal of Physiology* – Paris 104:5–18.

Chambon, P. 1995. The molecular and genetic dissection of the retinoid signaling pathway. *Recent Progress in Hormone Research* 50:317–332.

Chen, Y., and J. Jiang. 2013. Decoding the phosphorylation code in Hedgehog signal transduction. *Cell Research* 23(2): 186–200.

Chen, Y.F., N. Etheridge, and G.E. Schaller. 2005. Ethylene signal transduction. *Annals of Botany* 95:901–915.

Cheng et al., 2012. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Research* 22:1658–1667.

Chu, H.Y., et al. (2011). From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants, *Plant Journal* 66(1):66–79.

Chung, W.Y., S. Wadhawan, R. Szklarczyk, S.K. Pond, and A. Nukrutenko. 2007. A first look at ARFome: dual-coding genes in mammalian genomes. *PLoS Computational Biology* 3(5): e91.

Ciapponi, L., and G. Cenci. 2008. Telomere capping and cellular checkpoints: clues from fruit flies. *Cytogenetic and Genome Research* 122(3–4):365–373.

Cieśla, J., T. Frączyk, and W. Rode. 2011. Phosphorylation of basic amino acid residues in proteins: important but easily missed. *Acta Biochimica Polonica* 58(2): 137–147.

Claverys, J.P., M. Prudhomme, and B. Martin. 2006. Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annual Review of Microbiology* 60:451–475.

Cosgrove, M.S., and C. Wolberger. 2005. How does the histone code work? *Biochemistry and Cell Biology* 83:468–476.

Crick, F. 1970. Central dogma of molecular biology. *Nature*, 227:561–563.

Davidson, E.H. 2006. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. Academic Press, London, UK.

Davidson, E.H., and D.H. Erwin. 2006. Gene regulatory networks and the evolution of animal body plans. *Science* 311(5762):796–800.

Dawkins, Richard. 1996. *The Blind Watchmaker*. W. W. Norton & Company, Inc., London, UK.

de Lange, T. 2010. How shelterin solves the telomere end-protection problem. *Cold Spring Harbor Symposia on Quantitative Biology* 75:167–77.

Dill, K.A., S.B. Ozkan, M.S. Shell, and T.R. Weikl. 2008. The protein folding problem. *Annual Review of Biophysics* 37:289–316.

Dobrindt, U., B. Hochhut, U. Hentschel, and J. Hacker. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews Microbiology* 2:414–424.

Dufour, Y.S., P.J. Kiley, and T.J. Donohue. 2010. Reconstruction of the core and extended regulons of global transcription factors. *PLoS Genetics* 6(7): e1001027.

Dumesic, P.A., P. Natarajan, C. Chen, I.A. Drinnenberg, B.J. Schiller, J. Thompson, J.J. Moresco, J.R. Yates III, D.P. Bartel, and H.D. Madhani. 2013. Stalled spliceosomes signal for RNAi-mediated genome defense. *Cell* 152(5):957–68.

Emanuelsson, O. 2002. Predicting protein subcellular localisation from amino acid sequence information. *Briefings in Bioinformatics* 3(4):361–376.

Engelberg-Kulka, H., I. Yelin, and I. Kolodkin-Gal. 2009. Activation of a built-in bacterial programmed cell death system as a novel mechanism of action of some antibiotics. *Communicative & Integrative Biology* 2(3):211–212.

Fang, S.C., C. de los Reyes, and J.G. Umen. 2006. Cell size checkpoint control by the retinoblastoma tumor suppressor pathway. *PLoS Genetics* 2(10): e167.

Fell, D. 1997. *Understanding the Control of Metabolism*. Portland Press Ltd., London, UK.

Ficz, G. 2015. New insights into mechanisms that regulate DN methylation patterning. *The Journal of Experimental Biology* 218:14–20.

Flames, N., R. Pla, D.M. Gelman, J.L.R. Rubenstein, L. Puelles, and O. Marin. 2007. Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *The Journal of Neuroscience* 27(36): 9682–9695.

Forsberg, E.C., and S. Smith-Berdan. 2009. Parsing the niche code: the molecular mechanisms governing hematopoietic stem cell adhesion and differentiation. *Haematologica* 94(11):1477–1481.

Fuqua, C., M.R. Parsek, and E.P. Greenberg. 2001. Regulation of gene expression by cell-to-cell communication: Acyl-homoserine lactone quorum sensing. *Annual Review of Genetics* 35:439–468.

Gabdank, I., et al. 2010. FineStr: a web server for single-base-resolution nucleosome positioning. *Bioinformatics* 26:845–846.

Gabius, H.-J., S. André, J. Jiménez-Barbero, A. Romero, and D. Solís. 2011. From lectin structure to functional glycomics: principles of the sugar code. *Trends in Biochemical Sciences* 36(6):298–313.

Gatlin, L.L. 1972. *Information Theory and the Living System*. Columbia University Press, New York, NY.

Gazzaniga, M.S., R.B. Ivry, and G.R. Mangun. 2009. *Cognitive Neuroscience: The Biology of the Mind*, 3rd edition. W. W.

Norton & Company, Inc., New York, NY.

Gen New Highlights. 2015. Bacteria use CRISPR to recognize themselves. http://www.genengnews.com/gen-news-highlights/bacteria-use-crispr-to-recognize-themselves/81251146/

Gibbs, D.J., J. Bacardit, A. Bachmair, and M.J. Holdsworth. 2014. The eukaryotic N-end rule pathway: conserved mechanisms and diverse functions. *Trends in Cell Biology*, 24(10):603–611.

Gilbert, S.F. 2003. The reactive genome. In Müller, G.B., and S.A. Newmann (editors), *Origination of Organismal Form: Beyond the Gene in Development and Evolutionary Biology*, pp. 87–102. MIT Press, Cambridge, MA.

Gogarten, J.P., A.G. Senejani, O. Zhaxybayeva, L. Olendzenski, and E. Hilario. 2002. Inteins: structure, function, and evolution. *Annual Review of Microbiology* 56:263–287.

Gurdon, J.B., J.A. Byrne, S. Simonsson. 2003. Nuclear reprogramming and stem cell creation. *Proceedings of the National Academy of Sciences USA* 100 (Suppl 1):11819–11822.

Hall, R.M., and C.M. Collis. 1995. Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Molecular Microbiology* 15(4):593–600.

Harrow, J., A. Nagy, A. Reymond, T. Alioto, L. Patthy, S.E Antonarakis, and R. Guigó. 2009. Identifying protein-coding genes in genomic sequences. *Genome Biology* 10:201.

Hegde, R.S., and H.D. Bernstein. 2006. The surprising complexity of signal sequences. *Trends in Biochemical Sciences* 31(10):563–71.

Hofstadter, D.R. 1980. *Gödel, Escher, Bach: An Eternal Golden Braid*. Random House/Vintage Books, New York, NY.

Holoch, P.A., and T.S. Griffith. 2009. TNF-related apoptosis-inducing ligand (TRAIL): a new path to anti-cancer therapies. *European Journal of Pharmacology* 625(1–3):63–72.

Hou, Y.M., and P. Schimmel. 1988. A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* 333:140–145.

Huen, M.S., and J. Chen. 2010. Assembly of checkpoint and repair machineries at DNA damage sites. *Trends in Biochemical Sciences* 35(2):101–108.

Hughes, T. 2008. Cracking the second genetic code. *FASEB Journal* 22:262.2.

Hurst, L.D., C. Pál, and M.J. Lercher. 2004. The evolutionary dynamics of eukaryotic gene order. *Nature Reviews Genetics* 5:299–310.

Ibarra, B., et al., 2009. Proofreading dynamics of a processive DNA polymerase. *EMBO Journal* 28(18):2794–2802.

Ikegami, K., J. Ohgane, S. Tanaka, S. Yagi, and K. Shiota. 2009. Interplay between DNA methylation, histone modification and chromatin remodeling in stem cells and during development. *International Journal of Developmental Biology* 53:203–214.

Ishihama, A. 2000. Functional modulation of Escherichia coli RNA polymerase. *Annual Review of Microbiology* 54:499–518.

Ishikawa, K., H. Ishii, and T. Saito. 2006. DNA damage-dependent cell cycle checkpoints and genomic stability. *DNA and Cell Biology* 25:406–411.

Jády, B.E., P. Richard, E. Bertrand, and T. Kiss. 2006. Cell cycle-dependent recruitment of telomerase RNA and Cajal bodies to human telomeres. *Molecular Biology of the Cell* 17 (2):944–54.

Jaenisch, R., and A. Bird. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics* 33 (Suppl):245–254.

Janke, C. 2014. The tubulin code: molecular components, readout mechanisms, and functions. *Journal of Cell Biology* 206(4):461–472.

Jenuwein, T., and C.D. Allis. 2001. Translating the histone code. *Science* 293:1074–1080.

Jessell, T.M. 2000. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nature Genetics* 1:20–29.

Jolma, A., Y. Yin, K.R. Nitta, K. Dave, A. Popov, M. Taipale, M. Enge, T. Kivioja, E. Morgunova, and J. Taipale. 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* 527:384–388.

Jones, P.A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* 13(7):484–92.

Juven-Gershon, T., and J.T. Kadonaga. 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology* 339(2):225–229.

Kadyrova, L.Y., et al. 2013. A reversible histone h3 acetylation cooperates with mismatch repair and replicative polymerases in maintaining genome stability. *PLoS Genetics* 9:e1003899.

Keverne, E.B., D.W. Pfaff, and I. Tabansky. 2015. Epigenetic changes in the developing brain: effects on behavior. *Proceedings of the National Academy of Sciences* 112(22):6789–6795.

Kirschner, M.W., and J.C. Gerhart. 2005. *The Plausibility of Life: Resolving Darwin's Dilemma*. Yale University Press, New Haven, CT.

Klipp, E., W. Liebermeister, C. Wierling, A. Kowald, H. Lehrach, and R. Herwig. 2009. *Systems Biology*. Wiley-VCH Verlag GmbH & Co. KgaA, Weinheim, Germany.

Kolovos, P., T.A. Knoch, F.G. Grosveld, P.R. Cook, and A. Papantonis. 2012. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics & Chromatin* 5:1–8.

Komano, T. 1999. Shufflons: multiple inversion systems and integrons. *Annual Review of Genetics* 3:171–191.

Koonin, E.V., K.S. Makarova, and L. Aravind. 2001. Horizontal gene transfer in prokayrotes: quantification and classification. *Annual Review of Microbiology* 55:709–742.

Kosko, B., and S. Isaka. 1993. Fuzzy logic. *Scientific American* July:76–81.

Krysko, D.V., et al. 2005. Gap junctions and the propagation of cell survival and cell death signals. *Apoptosis* 10:459–469.

Kundaje, A. et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* 518:317–330.

Kunkel, T.A., and K. Bebenek. 2000. DNA replication fidelity. *Annual Review of Biochemistry* 69:497–529.

Kunkel, T.A., and D.A. Erie. 2005. DNA mismatch repair. *Annual Review of Biochemistry* 74:681–710.

Lazazzera, B.A. 2001. The intracellular function of extracellular signaling peptides. *Peptides* 22:1519–1527.

Lee, H.J., T.A. Hore, and W. Reik. 2014. Reprogramming the methylome: erasing memory and creating diversity *Cell Stem Cell* 14:710–719.

London, M., and M. Hausser. 2005. Dendritic computation. *Annual Review of Neuroscience* 28:503–532.

Markram, H., et al. 2015. Reconstruction and simulation of neocortical microcircuitry. *Cell* 163:456–492.

Marquardt, T., and S.L. Pfaff. 2001. Cracking the transcriptional code for cell specification in the neural tube. *Cell* 106:651–654.

Mattick, J.S., and I.V. Makunin. 2006. Noncoding RNA. *Human Molecular Genetics* 15(1):R17-R29.

Michalak, P. 2008. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* 91(3):243–248.

Miller, K.M., et al. 2010. Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining. *Nature Structural and Molecular Biology* 17:1144–1151.

Miwa, Y., A. Nakata, A. Ogiwara, M. Yamamoto, and Y. Fujita. 2000. Evaluation and characterization of catabolite-responsive elements (cre) of Bacillus subtilis. *Nucleic Acids Research* 28:1206–1210.

Modrich, P., and R. Lahue. 1996. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annual Review of Biochemistry* 65:101–133.

Moller-Krull, M., et al. 2008. Beyond DNA: RNA editing and steps toward Alu exonization in primates. *Journal of Molecular Biology* 382(3):601–599.

Montañez, G., R.J. Marks II, J. Fernandez, and J.C. Sanford. 2013. Multiple overlapping genetic codes profoundly reduce the probability of beneficial mutations. In Marks, R.J. II, M.J. Behe, W.A. Dembski, B.L. Gordon, and J.C. Sanford (editors), *Biological Information: New Perspectives,* pp. 139–167. World Scientific, Tuck Link, Singapore.

Müller, G.B., and S.A. Newmann (editors). 2003. *Origination of Organismal Form: Beyond the Gene in Development and Evolutionary Biology*. MIT Press, Cambridge, MA.

Murai, K.K., and E.B. Pasquale. 2004. Eph receptors, ephrins, and synaptic function. *Neuroscientist* 10:304–314.

Murphy, P.V., S. André, and H.-J. Gabius. 2013. The third dimension of reading the sugar code by lectins: design of glycoclusters with cyclic scaffolds as tools with the aim to define correlations between spatial presentation and activity. *Molecules* 18:4026–4053.

Musacchio, A. 2011. Spindle assembly checkpoint: the third decade. *Philosophical Transactions of the Royal Society of London B, Biological Sciences* 366(1584): 3595–604.

Neel, N.F., E. Schutyser, J. Sai, G.H. Fan, and A. Richmond. 2005. Chemokine receptor internalization and intracellular trafficking. *Cytokine and Growth Factor Reviews* 16:637–658.

Nezi, L., and A. Musacchio. 2009. Sister chromatid tension and the spindle assembly checkpoint. *Current Opinion in Cell Biology* 21:785–795.

Nguyen, V.C., et al. 2010. Replication stress checkpoint signaling controls tRNA gene transcription. *Nature Structural and Molecular Biology* 17(8):976–81.

Nicolelis, M.A.L., and S. Ribeiro. 2006. Seeking the neural code. *Scientific American* 295:70–77.

Nishikura, K. 2010. Functions and regulation of RNA editing by ADAR deaminases. *Annual Review of Biochemistry* 79:321–349.

Noble, D. 2015. Evolution beyond neo-Darwinism: a new conceptual framework. *Journal of Experimental Biology* 218:7–13.

Ochman, H., J.G. Lawrence, and E.A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304.

Osbourn, A.E., and B. Field. 2009. Operons. *Cellular and Molecular Life Sciences* 66(23):3755–3775.

Oyama, S. 2002. *The Ontogeny of Information: Developmental Systems and Evolution*. 2nd edition. Duke University Press, Durham, NC.

Page, A.M., and P. Hieter. 1999. The anaphase-promoting complex: new subunits and regulators. *Annual Review of Biochemistry* 68:583–609.

Paszkowski, J., and S.A. Whitham. 2001. Gene silencing and DNA methylation processes. *Current Opinion in Plant Biology* 4:123–129.

Pauli, A., J.L. Rinn, and A.F. Schier. 2011. Non-coding RNAs as regulators of embryogenesis. *Nature Reviews* 12:136–149.

Pelkmans, L. 2005. Viruses as probes for systems analysis of cellular signaling, cytoskeleton reorganization and endocytosis. *Current Opinion in Microbiology* 8:331–337.

Pollard, T.D., and G.G. Borisy. 2003. Cellular motility driven by assembly and disassembly of actin filaments. *Cell* 112:453–465.

Pool, M.R. 2005. Signal recognition particles in chloroplasts, bacteria, yeast and mammals (review). *Molecular Membrane Biology* 22:3–15.

Putnam, C.D., E.J. Jaehnig, and R.D. Kolodner. 2009. Perspectives on the DNA damage and replication checkpoint responses in Saccharomyces cerevisiaem. *DNA Repair* 8(9):974–82.

Ran, F.A., et al. 2015. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* 520(7546):186–191.

Readies, C., and M. Takeichi. 1996. Cadherine in the developing central nervous system: an adhesive code for segmental and functional subdivisions. *Developmental Biology* 180:413–423.

Rino, J., and M. Carmo-Fonseca. 2009. The spliceosome: a self-organized macromolecular machine in the nucleus? *Trends in Cell Biology* 19(8):375–84.

Sabelli, P.A., et al. 2013. Control of cell proliferation, endoreduplication, cell size, and cell death by the retinoblastoma-related pathway in maize endosperm. *Proceedings of the National Academy of Sciences USA* 110:E1827–1836.

Salih, F., et al. 2008. Epigenetic nucleosomes: Alu sequences and CG as nucleosome positioning element. *Journal of Biomolecular Structure and Dynamics* 26:9–16.

Schmidt, C.K., and S.P. Jackson. 2013. On your mark, get SET(D2), go! H3K36me3 primes DNA mismatch repair. *Cell* 153:513–515.

Schneider, R., and R. Grosschedl. 2007. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes and Development* 21:3027– 3043.

Schübeler, D. 2015. Function and information content of DNA methylation. *Nature* 517:321–326.

Scruton, R. 1996. *A Short History of Modern Philosophy*. Routledge, Abingdon-on-Thames, UK.

Searle, J.S., et al. 2011. Proteins in the nutrient-sensing and DNA damage checkpoint pathways cooperate to restrain mitotic progression following DNA damage. *PLoS Genetics* 7(7):e1002176.

Segal, E., et al. 2006. A genomic code for nucleosome positioning. *Nature* 442:772–778.

Segurado, M., and J.A. Tercero. 2009. The S-phase checkpoint: targeting the replication fork. *Biology of the Cell* 101:617–627.

Serganov, A., and D.J. Patel. 2007. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature Reviews Genetics* 8:776–790.

Shapiro, J.A. 2006. Genome informatics: the role of DNA in cellular computations. *Biological Theory* 1(3):288–301.

Shapiro, J.A. 2011. Cognitive aspects of genome function. Symposium on Neurobiological Correlates of Interpersonal Relations, Freiburg, Germany, October 15.

Shapiro, J.A. 2014. Physiology of the read-write genome. *Journal of Physiology* 592.11:2319–2341.

Shapiro, L., and D.R. Colman. 1999. The diversity of cadherins and implications for a synaptic adhesive code in the CNS. *Neuron* 23:427–430.

Shapiro, J.A., and R. von Sternberg. 2005. Why repetitive DNA is essential to genome function. *Biological Reviews* 80:1–24.

Sidiropoulou, K, E.K. Pissadaki, and P. Poirazi. 2006. Inside the brain of a neuron. *EMBO Reports* 7(9):886–892.

Signal Peptide Website: An Information Platform for Signal Sequences and Signal Peptides. http://www.signalpeptide.de/

Sonea, S., and L.G. Mathieu. 2001. Evolution of the genomic systems of prokaryotes and its momentous consequences. *International Microbiology* 4:67–71.

Song, J. 2007. EMT or apoptosis: a decision for TGF-beta. *Cell Research* 17(4): 289–290.

Spencer, P.S., E. Siller, J.F. Anderson, and J.M. Barral. 2012. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *Journal of Molecular Biology* 422:328–335.

Stangroom, J., and J. Garvey. 2005. *The Great Philosophers: From Socrates to Foucault*. Arcturus Publishing Limited, Longdon, UK.

Stanton, T.B. 2007. Prophage-like gene transfer agents—novel mechanisms of gene exchange for Methanococcus, Desulfovibrio, Brachyspira, and Rhodobacter species. *Anaerobe* 13:43–49.

Storz, G., Altuvia, S., and K.M. Wassarman. 2005. An abundance of RNA regulators. *Annual Review of Biochemistry* 74:199–217.

Strahl, B.D., and C.D. Allis. 2000. The language of covalent histone modifications. *Nature* 403:41–45.

Striedter, G.F. 2003. Epigenesis and evolution of brains: from embryonic divisions to functional systems. In Müller, G.B., and S.A. Newmann (editors), *Origination of Organismal Form: Beyond the Gene in Development and Evolutionary Biology*, pp. 287–304. MIT Press, Cambridge, MA.

Stuart, L.M., R.A. Ezekowitz. 2005. Phagocytosis: elegant complexity. *Immunity* 22:539–550.

Sugiyama, T., et al. 2005. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proceedings of the National Academy of Sciences USA* 102(1): 152–157.

Sullivan, B.A., M.D. Blower, and G.H. Karpen. 2001. Determining centromere identity: cyclical stories and forking paths. *Nature Reviews Genetics* 2:584–596.

Taft, R.J., K.C. Pang, T.C. Mercer, M. Dinger, and J.S. Mattick. 2010. Non-coding RNAs: regulators of disease. *Journal of Pathology* 220:126–139.

Takada, Y., X. Ye, and S. Simon. 2007. The integrins. *Genome Biology* 8(5):215.

Tam, C.K.P., J. Hackett, and C. Morris. 2005. Rate of inversion of the Salmonella enterica shufflon regulates expression of invertible DNA. *Infection and Immunity* 73(9):5568–5577.

Tejedor, J.R., and J. Valcárcel. 2010. Breaking the second genetic code. *Nature* 465(6):45–46.

Tentner, A.R., et al. 2012. Combined experimental and computational analysis of DNA damage signaling reveals context-dependent roles for Erk in apoptosis and G1/S arrest after genotoxic stress. *Molecular Systems Biology* 8:568.

Thanbichler, M., and L. Shapiro. 2008. Getting organized–how bacterial cells move proteins and DNA. *Nature Reviews Microbiology* 6:28–40.

Thomas, C.M., and K.M. Nielsen. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Review Microbiology* 3:711–721.

Togneri, R., and C.J.S. deSilva. 2003. *Fundamentals of Information Theory and*

Coding Design. Chapman & Hall/CRC, Boca Raton, FL.

Tomkins, M.G. 1975. The metabolic code. *Science* 189:760–763.

Trifonov, E.N. 1980. Sequence-dependent deformational anisotropy of chromatin DNA. *Nucleic Acids Research* 8:4041–4053.

Trifonov, E.N. 1981. Structure of DNA in chromatin. In Schweiger, H. (editor), *International Cell Biology 1980–1981*, pp. 128–138. Springer, Berlin, Germany.

Trifonov, E.N. 2011. Thirty years of multiple sequence codes. *Genomics Proteomics Bioinformatics* 9(1–2):1–6.

Trotta, C.R., F. Miao, E.A. Arn, S.W. Stevens, C.K. Ho, R. Rauhut, and J.N. Abelson. 1997. The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell* 89(6):849–858.

Truman, R. 2012. An evaluation of codes more compact than the natural genetic code. *Journal of Creation* 26(2):88–99.

Truman, R. 2012a. Information theory—part 1—overview of key ideas. *Journal of Creation* 26(3):101–106. http://creation.com/cis-1.

Truman, R. 2012b. Information theory—part 2: weaknesses in current conceptual frameworks. *Journal of Creation* 26(3):107–114. http://creation.com/cis-2.

Truman, R. 2012c. Information theory—part 3: introduction to coded information systems. *Journal of Creation* 26(3):115–119. http://creation.com/cis-3.

Truman, R. 2013. Information theory—part 4: fundamental theorems of coded information systems theory. *Journal of Creation* 27(1):71–77. http://creation.com/cis-4.

Truman, R. 2015. Nylon-eating bacteria—part 4: interpretation according to coded information system theory. *Journal of Creation* 29(3):46–52.

van de Lagemaat, L.N., L. Gagnier, P. Medstrand, and D.L. Mager. 2005. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. *Genome Research* 15(9):1243–1249.

Varshavsky A. 2011. The N-end rule pathway and regulation by proteolysis. *Protein Science* 8:1298–345.

Verdel, A., et al. 2009. Common themes in siRNA-mediated epigenetic silencing pathways. *International Journal of Developmental Biology* 53(2–3):245–257.

Verhey, K.J., and J. Gaertig. 2007. The tubulin code. *Cell Cycle* 6:(17):2152–2160.

Wagner, S., et al. 2015. The EF-hand Ca2 binding protein MICU choreographs mitochondrial Ca2 dynamics in arabidopsis. *The Plant Cell* 27(11):3190–3212.

Walhout, M., M. Vidal and J. Dekker (editors). 2013. *Handbook of Systems Biology: Concepts and Insights*. Elsevier, London, UK.

Walsh, C.M., and A.L. Edinger. 2010. The complex interplay between autophagy, apoptosis, and necrotic signals promotes T-cell homeostasis. *Immunological Reviews* 236:95–109.

Wellik, D.M. 2007. Hox patterning of the vertebrate axial skeleton. *Development Dynamcs* 236:2454–2463.

Widelitz, R. 2005. Wnt signaling through canonical and non-canonical pathways: recent progress. *Growth Factors* 23:111–116.

Wikipedia. n.d. Von Neumann architecture. https://en.wikipedia.org/wiki/Von_Neumann_architecture.

Williams, T.L., D.L. Levy, S. Maki-Yonekura, K. Yonekura, and E.H. Blackburn. 2010. Characterization of the yeast telomere nucleoprotein core. Rap1 binds independently to each recognition site. *The Journal of Biological Chemistry* 285:35814–35824.

Willmer, P. 2003. Convergence and homplasmy in the evolution of organizmal form. In Müller, G.B., and S.A. Newmann (editors), *Origination of Organismal Form: Beyond the Gene in Development and Evolutionary Biology*, pp. 33–50. MIT Press, Cambridge, MA.

Woodward, T.E., and J.P. Gills. 2012. *The Mysterious Epigenome: What Lies Beyond DNA*. Kregel Publications, Grand Rapids, MI.

Wulfing C., I. Tskvitaria-Fuller, N. Burroughs, M.D. Sjaastad, J. Klem, J.D. Schatzle. 2002. Interface accumulation of receptor/ligand couples in lymphocyte activation: Methods, mechanisms, and significance. *Immunological Reviews* 189:64–83.

Yamada, S., and W.J. Nelson. 2007. Synapses: Sites of cell recognition, adhesion, and functional specification. *Annual Review of Biochemistry* 76:267–294.

Young, E. 2001. Packaging proteins may be second genetic code. *New Scientist*, August 9. https://www.newscientist.com/article/dn1140-packaging-proteins-may-be-second-genetic-code/

Yuh, C.H., H. Bolouri, and E.H. Davidson. 1998. Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* 279:1896–1902.

Zeiler, M., W.L. Straube, E. Lundberg, M. Uhlen, and M. Mann. 2012. A protein epitope signature tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Molecular & Cellular Proteomics* 11(3): O111.009613.

Zetsche, B., et al. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-cas system. *Cell* 163:759–771.

Zhou, T., N. Shen, L. Yang, N. Abe, J. Horton, R.S. Mann, H.J. Bussemaker, R. Gordân, and R. Rohs. 2015. Quantitative modeling of transcription factor binding specificities using DNA shape. *Proceedings of the National Academy of Sciences USA* 112(15):4654–4659.

Creation Research Society Quarterly 2016. 52:309–324.

Citrate Utilizing Mutants of *Escherichia coli*

Kevin Anderson*

Abstract

As part of a "long-term evolution experiment," populations of *Escherichia coli* have been grown for thousands of generations in a consistent environment. During this experimental period, various mutations have altered the bacteria's phenotype. Some of these phenotypic changes have included larger cell size and faster growth rates. The wild-type strain of *E. coli* can use citrate as an energy source in anaerobic conditions, but not in aerobic conditions. However, after 31,500 generations, a population of mutants developed that could aerobically utilize citrate. The formation of these Cit$^+$ mutants entails an intriguing series of mutational steps involving both the citrate operon and other metabolic related genes. Thus, this new phenotype is frequently identified as (1) an example of the "birth of new genes," and (2) how random mutation and natural selection can drive neo-Darwinian evolution. However, all the mutations detected in the Cit$^+$ phenotype involve rearrangement of preexisting genes, loss of preexisting gene expression, or loss of preexisting regulation. Thus, the Cit$^+$ mutants fail to provide a genetic example of the origin of new genes or regulatory systems. In contrast, these mutants fit precisely within predictions of a creation model; organisms have a programmed ability to adapt to specific environmental conditions.

Introduction

Both the neo-Darwinian evolution model and the biblical creation model predict that organisms will adapt to their environment. However, they make distinctly different predictions regarding the manner of this adaptation. These differing predictions can be useful in evaluating the accuracy and scientific validity of each model.

The creation model proposes that life arose from the direct action of a creator. Organisms do not share a common ancestry with all other organisms. For example, elephants were created as the elephant "kind," and do not share a common ancestry with trees, fish, or primates. DNA mutations and other genetic events can alter an organism's phenotype, but such changes are either from mutations (or other forms of DNA damage) or part of the programmed ability of the organism to adapt.

Thus, the creation model predicts that organisms were created with the capacity to undergo adaptive modifications that can help them survive environ-

* Kevin Anderson, Ph.D., Van Andel Creation Research Center, Creation Research Society, 6801 N Hwy 89, Chino Valley, AZ 86323, anderson@crsvarc.com

Accepted for publication April 22, 2016

mental changes of climate, nutrition, etc. Such modifications can involve physiological adjustments, such as increasing or decreasing cellular levels of specific proteins. These modifications may also involve genetic alterations, which can include mutations, horizontal gene transfer, and epigenetic events.

There are also limits to the extent of these programmed adaptations. Bacterial antibiotic resistance is a classic example. Individual populations of bacteria develop resistance to specific antibiotics by mutation, gene transfer, and possibly epigenetic factors. In the presence of the antibiotic, this resistance gives a clear adaptive advantage over populations that lack resistance. Yet, gene transfer merely moves preexisting genetic systems into the cell, and mutations that provide resistance are generally degenerative (Anderson, 2005).

In contrast, the neo-Darwinian evolution model predicts all life on Earth shares a common ancestry (i.e., universal common descent). Contemporary life slowly originated and transformed from lower forms of life over immense periods of time. As such, vast physiological and genetic changes have occurred during biological history, leading to the development of new functional systems (e.g., wings, legs, eyes, and brains).

Such changes are required by neo-Darwinism to account for universal common descent. Organisms must not only have a mechanism to adapt to differing environmental conditions (as also predicted by the creation model) but also have an additional mechanism to develop new genetic systems that they did not previously possess. This mechanism must account for the "birth" of new genes, new regulatory controls, new proteins, and new transport systems. Only with such a mechanism can invertebrates become vertebrates, flightless creatures acquire flight, and marine creatures acclimate to an air-breathing physiology. Thus, a key difference between the two models is that creation predicts only limited changes, whereas neo-Darwinian evolution requires almost unlimited levels of change.

Proponents claim that the mechanism for this Darwinian transformation is random mutations, which alter an organism's phenotype (Carlin, 2011; Merlin, 2010). If the alteration provides an adaptive advantage, natural selection will promote its spread within the population (Patterson, 1978; Merlin, 2010). Dawkins (1996, p. 79) refers to this as "the non-random survival of random variants."

Ruiz-Orera et al. (2015) note that "the birth of new genes is an important motor of evolutionary innovation." Yet the search for examples of random mutations that give birth to these new genes has proven rather difficult (Anderson and Purdom, 2008; Depew and Weber, 2011; Margulis and Sagan, 2002; Noble, 2015). In fact, historically, Darwinists have generally focused on phenotypic changes with little attention to the resulting genotype. If the phenotype is positively selected, then it is assumed to be an example of the necessary "gene birth" (e.g., Dawkins, 1996; Coyne, 2009; Zimmer, 2001).

However, phenotypic advantage for an organism can frequently result from genetic degeneration. Some classic examples would be loss of transport proteins that provide antibiotic resistance to bacteria (Anderson, 2005), malaria resistance by improper folding of hemoglobin (Cholera et al., 2008), and HIV resistance due to loss of the CCR5 protein (Allers et al., 2011). Each of these phenotypes has been offered as an example of "evolution in action," yet all directly result from mutations that reduce or eliminate preexisting functional systems.

This illustrates a key difference between genotype and phenotype of an organism. Mutations that provide a "beneficial" phenotype can result from a degenerative genotype. These types of mutations cannot account for the origin of the CCR5 protein, hemoglobin, transport systems, etc.—only their degeneration. This is fully consistent with a creation model; proteins and regulatory systems were created fully formed and changes are limited. In contrast, these do not provide an example of the required neo-Darwinian mechanism for the origin of genetic systems, nor the virtually unlimited potential of biological change.

In search of examples for this genetic mechanism, a popular study frequently cited by neo-Darwinists is the "long-term evolution experiment" (LTEE) using *Escherichia coli*. Results of this experiment provide a fascinating variety of different mutants with various phenotypic adaptations, including the development of a new citrate utilizing phenotype. Results from the LTEE have also given rise to various claims about the adaptive capabilities of bacteria to generate new genes and functions. However, the criterion for these conclusions is primarily a positively selected phenotype, while genotypic changes are frequently a secondary consideration.

Long-Term Evolution Experiment

In 1988 Richard Lenski began an interesting study of bacterial adaptation. Using a culture of *E. coli* B, substrain REL606 or REL607 (which are designated as the wild-type strains), the bacterium was cultivated aerobically in growth medium that contained a limited amount of glucose (which served as the sole energy source) (Lenski, 2010). The organism was incubated in 12 separate flasks (six containing REL606 and six containing REL607) at a consistent 37ºC. Cultures in each flask were allowed to grow for 24 hours (approximately 6.6 generations), and 1% of each culture was transferred to 100 ml of fresh media for another 24 hours. Every 500 generations, a sample from each flask was used for detailed analysis

of both the phenotype and genotype of the bacterium.

Because of the limited glucose in the medium, the wild-type strain achieved only modest growth during each 24-hour period. Over time, subpopulations of the bacterium began to adapt to the medium. After 20,000 generations, some mutants had developed a larger cell size and grew approximately 70% faster than the wild-type strain (Lenski and Travisano, 1994; Novak et al., 2006). However, Lenski (2003) estimates that despite millions of mutations occurring in the bacteria during the first 20,000 generations, less than 100 point mutations and no more than 20 "beneficial" mutations became fixed within the population.

While some mutants grow significantly faster than the wild-type at the experimental temperature (37°), they generally have a reduced growth rate at lower (20°) and higher (42°) temperatures (Cooper et al., 2001a) and a reduced cold-stress tolerance (Sleight et al., 2006). These mutants also have a decreased survival in prolonged stationary phase. Those mutants with enlarged cell sizes show an increased susceptibility to osmotic pressure (Philippe et al, 2009). Some of these faster growing mutants lost MutT activity, which dramatically increases their mutation rate, especially mutations causing a transversion (purine → pyrimidine conversions) (Barrick et al., 2009). In addition, some of these mutants have a reduced level of ribose operon activity (Cooper et al., 2001b), maltose regulon activity (Pelosi et al., 2006), or flagella activity (Cooper et al., 2003).

This trade-off of features has been termed "antagonistic pleiotropy," where an adaptive advantage is gained by losing a function or system that is not essential for the current environment (Cooper and Lenski, 2000; Ostrowski et al., 2005). Interestingly, in the absence of examples showing "the birth of new genes" by random mutations, evolutionists have frequently pointed to antagonistic pleiotropy as a driving mechanism for evolutionary innovation (e.g., Rose, 1985; Fry, 1993; Olson, 1999; Roff and Fairborn, 2006). This phenomenon is even identified as the primary contributor to mutational adaptations during the LTEE (Cooper and Lenski, 2000).

However, antagonist pleiotropy is a degenerative event. Preexisting systems are lost, even if that loss provides a temporal or conditional advantage for the organism. While antagonistic pleiotropy fits within the predictions of both the neo-Darwinian and creation models, it fails to offer a genetic mechanism for the origin of the lost systems. Thus, it does not provide neo-Darwinian evolution the needed genetic mechanism for universal common descent.

What is more, identifying the formation of a "new function," "new system," "new gene," etc. is very context dependent. Is it a "new function" because the organism now metabolizes a unique substrate it could not previously utilize, or is it simply that a preexisting enzyme has a reduced specificity enabling it to bind with a larger pool of substrates? Is it a "new system" because the organism has a different growth characteristic, or did loss of one or more transport proteins alter its physiological balance? Did the organism form a "new gene" that it did not previously possess, or is it merely expressing a "silent" gene because of loss of a regulatory protein? Do these represent evolution of new features or simply loss of preexisting systems? Making these distinctions illustrates the importance of identifying changes to the genotype and not just the phenotype.

Citrate Mutants

As a chelating agent, a low concentration of citrate was added to the base growth medium used during the LTEE (Lenski, 2010). The wild-type strain (*E. coli* B) is not able to utilize citrate as an energy source (Cit^-) in an aerobic environment (Scheutz and Strockbine, 2005). However, after 31,500 generations, Cit^+ mutants were detected that could aerobically utilize citrate (Blount et al., 2008). Since the primary energy source (glucose) was limited in the medium, mutants that could also utilize citrate for additional energy possess a distinct growth advantage versus the wild-type strain.

These Cit^+ mutants have become a popular example of "evolutionary innovation" by mutation and selection, and they certainly provide a unique insight into the adaptive capacity of bacteria. In fact, a standard classification characteristic of *E. coli* is that it cannot utilize citrate in aerobic conditions (Scheutz and Strockbine, 2005). Thus, by this definition, the Cit^+ mutants are no longer *E. coli*. While Lenski and coworkers are not yet claiming a new species of *Escherichia* has "evolved" (Blount et al., 2008), they point to the Cit^+ mutants as examples of how speciation can occur. Not surprisingly, this makes these mutants of special interest to both creationists and evolutionists.

Citrate Operon

E. coli can utilize citrate as an energy source, but only in anaerobic conditions. The citrate operon of *E. coli* B is comprised of several genes (Figure 1). A characteristic of bacterial operons is that transcription of multiple genes is initiated by the same promoter, resulting in a polycistronic transcript (i.e., mRNA with multiple open reading frames). In the case of the *cit* operon (*citCDEFXGT*), the promoter is adjacent to *cit*C. When citrate is available to the cell, this promoter is activated by the CitA-CitB signal transduction system (Yamamoto et al., 2008). This transduction system is the product of *cit*A and *citB* located upstream of the citrate operon (Figure 1). CitA is a membrane protein that apparently is inactivated by an increase of the redox potential within the cell (Yamamoto et al., 2009). This inactivation prevents CitA-CitB from

Figure 1. Citrate operon and upstream genes in *E. coli* B. P_1 and P_2 denote promoter locations, and arrows indicate direction of transcription. The P_2 promoter is activated in anaerobic conditions, but remains inactive in aerobic conditions. (Adapted from the *E. coli* B str. REL606 genome sequence provided by www.paricbrc.org.)

functioning under aerobic conditions, so the citrate operon is not expressed (Scheu et al., 2012).

CitT functions as a transport protein for citrate, specifically serving as a citrate/succinate antiporter (Pos et al., 1998) (Figure 2). The lack of CitA-CitB activity under aerobic conditions prevents expression of CitT. Without this protein, *E. coli* cannot transport citrate into the cell, which results in the aerobic Cit^- phenotype.

Interestingly, when *citT* was placed under direct control of a plasmid promoter (e.g., insertion into pUC19), expression of the gene occurs aerobically (Pos et al., 1998). *E. coli* transformed with this plasmid is Cit^+ in aerobic conditions. Thus, the lack of CitT appears to be the primary limiting factor for aerobic utilization of citrate. Many of the other *cit* operon genes are involved in anaerobic utilization of citrate and thus are not necessary for the aerobic Cit^+ phenotype. The study of Pos et al. (1998) demonstrates that *E. coli* already possesses metabolic pathways for aerobic catabolism of citrate; the organism just lacks an aerobic transport system.

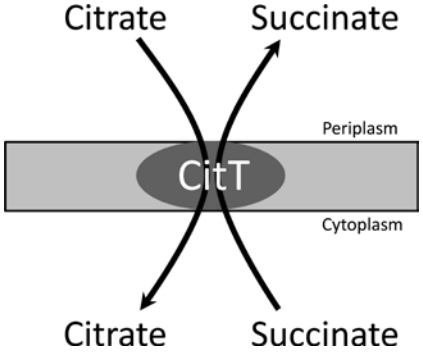


Figure 2. Antiporter activity of CitT showing simultaneous import of citrate and export of succinate.

Mutant Genotype

The initial report of the Cit^+ mutants contains little genotypic information (Blount, et al. 2008). Rather, the primary focus of the report was the phenotypic characteristics of the mutants: time of appearance in the population, growth characteristics, and frequency of reoccurrence. Using stored isolates from earlier generations, the researchers attempted to retrace the development of the Cit^+ phenotype. They observed that Cit^+ is more likely to reappear in a population after 20,000 generations than from populations of earlier generations. Blount et al. (2008) suggest this indicates a potentiating mutation occurred about 10,000 generations before the first Cit^+ phenotype was detected (making subsequent development of the Cit^+ mutants more likely).

Since the total number of predicted mutations during the first 30,000 generations is far greater than the genome size of the *E. coli*, this would suggest that the Cit^+ phenotype did not result from just a single point mutation. After a more complete genetic analysis, Blount et al. (2012) conclude that one or more initial mutations (at or after 20,000 generations) are required before the Cit^+ phenotype develops. From this they deduce that the formation of the phenotype involves three steps: potentiation, actualization, and refinement. The potentiation step establishes a physiological background that enables the aerobic Cit^+ phenotype to ultimately develop. The actualization step marks the occurrence of a weak Cit^+ phenotype, and the refinement step enables the mutant to more proficiently utilize citrate.

A number of different mutations were found in the Cit^+ population, as well as the population that gave rise to the Cit^+ phenotype. Many of these mutations appear to be unrelated to citrate utilization. However, several have been identified as either directly or indirectly related to the development of the Cit^+ mutants (Table 1).

Potentiation

As a possible potentiation step, Quandt et al. (2015) identified several mutations that appear in the *E. coli* population at 25,000 generations. In particular, they found that a point mutation of *gltA* may help establish a favorable physiological background for the subsequent development of the Cit^+ phenotype. Expression of *gltA* produces citrate synthase. This enzyme catalyzes the first step of the Krebs cycle: condensation of oxaloacetate and acetyl-CoA. Citrate synthase is inhibited by NADH. As NADH levels increase, the synthase becomes less active. This inhibition helps the cell maintain its redox homeostasis. As the level of reduced electron carries increases, Krebs

Step	Gene	Mutation	Phenotype of Mutation
Potentiation:			
	gltA	*gltA1*:	
		reduce NADH inhibition of enzyme	increase citrate synthase activity
	iclR	reduce activity of regulatory protein	increase expression of glyoxylate pathway enzymes
	arcB	reduce activity of regulatory protein	increase expression of Krebs cycle and glyoxylate pathway enzymes
Actualization:			
	citT	Promoter capture:	
		IS3 insertion into *citG*	aerobic CitT activity
		citT downstream of *fimB*	aerobic CitT activity
		rnk/*citG*/*citT* fusion	aerobic CitT activity
Refinement:			
	dctA	Overexpression of *dctA*	aerobic transport of high levels of succinate
	gltA	*gltA2*:	
		increase enzyme K_m value	reduce citrate synthase activity
		gltA2-R:	
		reduce gene transcription	reduce citrate synthase activity
	aceB	truncated transcript	reduce isocitrate lyase activity

Table 1. Genes identified in the potentiation, actualization, and refinement steps during the development of the *E. coli* Cit$^+$ phenotype.

cycle activity slows, decreasing oxidative release of electrons.

E. coli can produce acetate as an electron sink during glucose metabolism (Wolfe, 2005). Once the medium's limited glucose supply is depleted, the bacterium can utilize this acetate for additional energy. A point mutation (identified as *gltA1*) reduces the inhibitory effect of NADH, uncoupling the Krebs cycle from cellular levels of reduced electron carriers. As a result, the cell can achieve an increase of citrate synthase activity, which drives greater conversion of acetate to citrate. This would help maximize acetate catabolism after the glucose is consumed.

In addition, mutations to *iclR* and *arcB* also may contribute to a potentiating background (Quandt et al., 2015). Both IclR and ArcB are regulatory proteins that reduce expression of enzymes involved in the Krebs cycle and the glyoxylate shunt pathway. This shunt is an important pathway for acetate metabolism, as it enables acetyl-CoA to be converted to important anabolic intermediates (e.g., malate and oxaloacetate) rather than be lost by decarboxylation in the Krebs cycle (Cronan and Laporte, 2006) (Figure 3). The detected mutations of *iclR* and *arcB* reduce the inhibitory effect of IclR and ArcB, enabling increased production of Krebs cycle and glyoxylate pathway enzymes (Quandt et al., 2015). As with the *gltA1* mutation, these mutations may potentially increase acetate utilization and also reduce acetate excretion.

While evidence suggests *gltA1*, and perhaps to a lesser extent the *iclR* and *arcB* mutations, have a potentiating role in development of the Cit$^+$ phenotype, their exact contribution to this phenotype is still not clear (Quandt et al., 2015). Acting together, these mutations increase acetate metabolism by increasing activity of the Krebs cycle and the glyoxylate pathway (Figure 3). Presumably, this helps increase the intracellular level of succinate or other C_4-dicarboxylates that can be a product of acetate metabolism. Since these molecules are exported in exchange for citrate (Figure 2), their increased cellular concentration may be a beneficial first step for increased citrate transport.

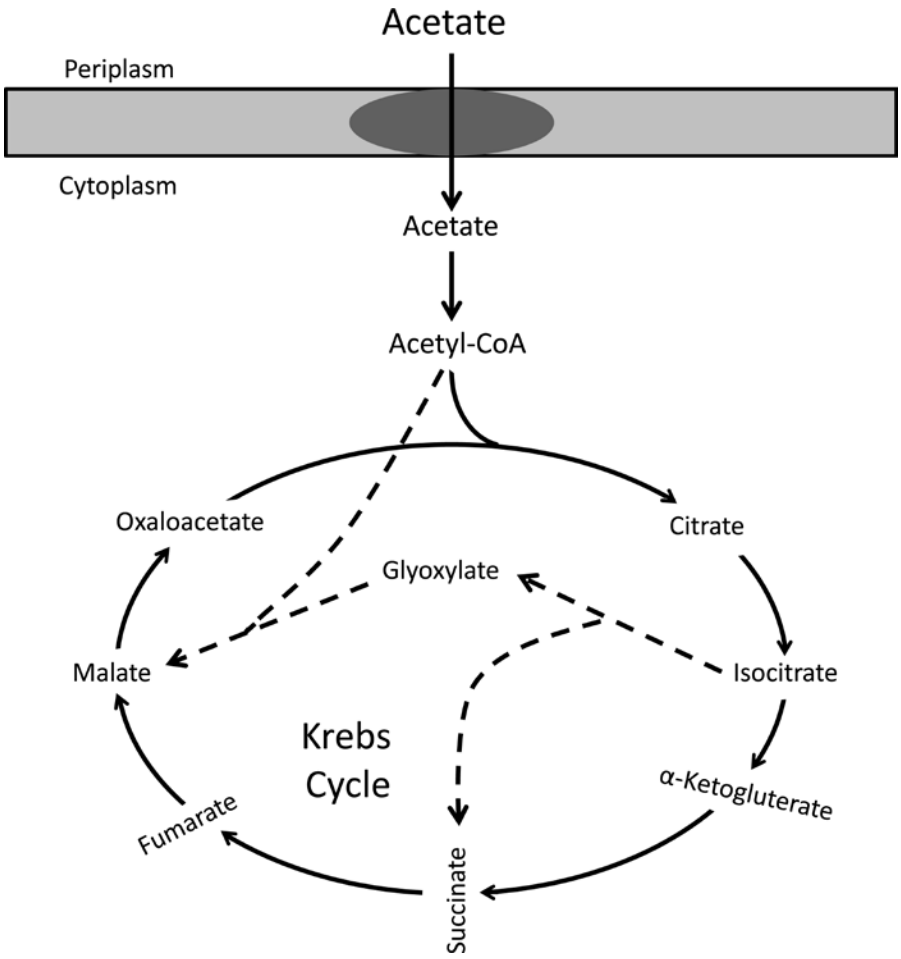


Figure 3. Pathway of acetate metabolism. Several potentiating mutations apparently drive greater activity of the Krebs cycle and the glyoxylate shunt pathway (shown with dashed lines).

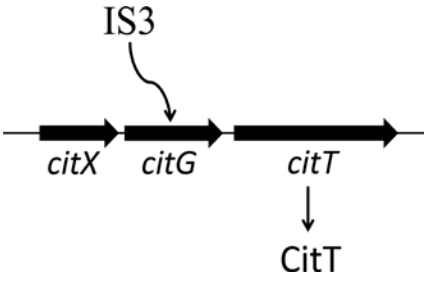


Figure 4. Insertion of an IS3 element into different locations of *citG*. The IS3 promoter enables aerobic expression of the downstream *citT*.

Interestingly, when the Cit+ phenotype develops, these potentiating mutations likely become detrimental (Quandt et al., 2015). When citrate is the sole carbon and energy source, mutations increasing acetate utilization no longer provide a benefit. These mutations may cause compounds (such as acetyl-CoA and oxaloacetate) to be unnecessarily catabolized rather than serve as needed anabolic intermediates. These mutations may also make it more difficult for the Cit+ mutants to maintain a redox homeostasis when catabolizing citrate.

Actualization

Within the variety of mutations found in the Cit+ population, only two appear to be essential for the full Cit+ phenotype (Quandt et al., 2014). One of these mutations is a "promoter capture" event. This capture involves placing *citT* under the control of an alternate promoter. Since such captures were detected in all Cit+ mutants studied, this suggests it is a key mutation for aerobic citrate utilization. However, this mutation only results in a weak Cit+ phenotype. Thus, it is possibly the actualization step proposed by Blount et al. (2012).

One form of promoter capture results from the insertion of the IS3 element into different sites of *citG*. Since IS3 possess a promoter that can activate adjacent genes (Charlier et al., 1982), this insertion initiates transcription of the downstream *citT* (Figure 4). The IS3 promoter is not under aerobic/anaerobic regulation; thereby it enables expression of *citT* in aerobic conditions.

A single Cit+ mutant was found to have an inversion that moved the *citG*/*citT* segment downstream of *fimB* (Blount et al., 2012) (Figure 5). This enables the *fimB* promoter to initiate expression of *citT*. As with the IS3 promoter, the *fimB* promoter is not under aerobic/anaerobic control, so *citT* expression can occur aerobically.

All Cit+ mutants that were analyzed also possess a rearrangement bringing the *citG*/*citT*/*rna* region downstream of the *rnk* gene (Blount et al., 2012) (Figure 6A). This rearrangement results in a portion of *citG* aligning with a segment of *rnk*. This forms a *rnk*/*citGT* fusion, enabling *citT* expression to be driven by the *rnk* promoter (Figure 6B). The *rnk* promoter enables a low expression of *citT* in aerobic conditions, resulting in a weak Cit+ phenotype (Blount et al, 2012).

In addition, several Cit+ mutants possess tandem duplications of the *rnk*/*citGT* fusion (Figure 6C). An increased number of fusion copies appear to

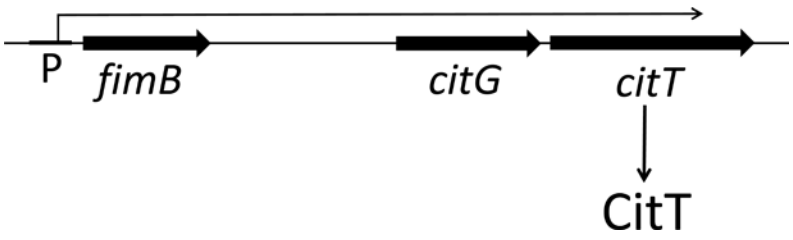


Figure 5. Chromosome rearrangement moves *citG* and *citT* downstream of *fimB*. This allows expression of *citT* via the *fimB* promoter.

increase the production of the CitT transporter protein (Blount et al., 2012). This results in a stronger Cit$^+$ phenotype. Potentially, the increase in cellular levels of this transporter (1) increases internal levels of citrate, driving greater citrate metabolism, and (2) enables the cell to more effectively scavenge the low levels of citrate in the medium.

Refinement

As mentioned above, the CitT transporter functions as an antiporter; exchanging citrate for C_4-dicarboxylates (such as succinate or fumarate) (Pos et al., 1998). In this capacity, CitT transports citrate into the cell and simultaneously exports succinate (Figure 2). Inadequate levels of a C_4-dicarboxylate will reduce the activity of the CitT antiporter, decreasing citrate transport.

dctA Mutation

Quandt et al. (2014) observe that the other mutation necessary for the full Cit$^+$ phenotype involves *dctA*. This gene expresses a protein that transports succinate and other C_4-dicarboxylates into the cell (Golby et al., 1999; Steinmetz et al., 2014) (Figure 7). *E. coli* has several anaerobic transport systems for C_4-dicarboxylates, but only DctA appears to serve as an aerobic transporter (Davies et al., 1999). DctA cotransports succinate and protons (H^+), linking it with the proton motive force of the organism.

In the Cit$^+$ mutants, a mutation is located 20 bases upstream of the *dctA* open reading frame, suggesting it falls within the promoter/operator region of the gene. This mutation causes overexpression of *dctA* (Quandt et al., 2014), giving the cell a higher copy number of the DctA transporter. In turn, increased numbers of DctA can increase the level of imported succinate, providing more succinate for the CitT antiporter (Figure 8).

The *dctA* mutation alone does not yield a Cit$^+$ phenotype and may actually be deleterious to the bacterium (Quandt et al., 2014). However, without the overproduction of DctA, the mutants do not achieve the full Cit$^+$ phenotype. Apparently without excess DctA, insufficient succinate is available to drive full activity of CitT. Thus, the *dctA* mutation is probably a "refinement." Loss of the *dctA* mutation in Cit$^+$ mutants reduces their level of citrate utilization (Quandt et al., 2014).

Because *E. coli* can oxidize citrate to succinate (Figure 9), each mole of transported citrate could result in the production of an equal mole of succinate for cotransporting by CitT. By this scenario, the role of DctA (esp. its overproduction) would appear unnecessary and certainly not a key component for the full Cit$^+$ phenotype. Rather, succinate production from the oxidation of citrate should be sufficient for full CitT activity.

However, it is unlikely that *E. coli* makes equal moles of succinate per mole of citrate. Since no amino acids and only one vitamin (thiamine) were added to the growth medium of the LTEE, *E. coli* must biosynthesize these compounds. Using various metabolic intermediates, *E. coli* can synthesize different amino acids and vitamins (e.g., α-ketoglutarate serves as the anabolic intermediate precursor for biosynthesis of glutamate and proline) (Lengeler et al., 1999). Since *E. coli* does not normally catabolize citrate under aerobic conditions, some details of this metabolism currently remain unknown.

The tandem duplicates of the *rnk*/*citGT* fusion enable overproduction of CitT. High copy numbers of CitT provide increased importing of citrate, even when the citrate concentration in the medium is low. However, increased CitT activity requires more C_4-dicarboxylate for cotransporting. Hence, increased activity of *dctA* assures that sufficient levels of succinate are available to drive full activity of the multiple copies of CitT (Figure 8). The overproduction of DctA fits within this physiological setting, and explains the results observed by Quandt et al. (2014). Correspondingly, the overproduction of CitT balances the excess amounts of succinate being imported by the high levels of DctA.

Thus, the two mutations appear to have an epistatic relationship. Excess production of DctA has no benefit without being counterbalanced by the excess production of CitT. Also, without sufficient succinate for cotransporting, excess production of CitT only gives a weak Cit$^+$ phenotype. Hence, overproduction of both CitT and DctA is necessary for the full Cit$^+$ phenotype, while overproduction of one transporter without overproduction of both has far less (if any) benefit for the bacterium.

Secondary Mutations

While not necessary for the full Cit$^+$ phenotype (which only requires the *citT*

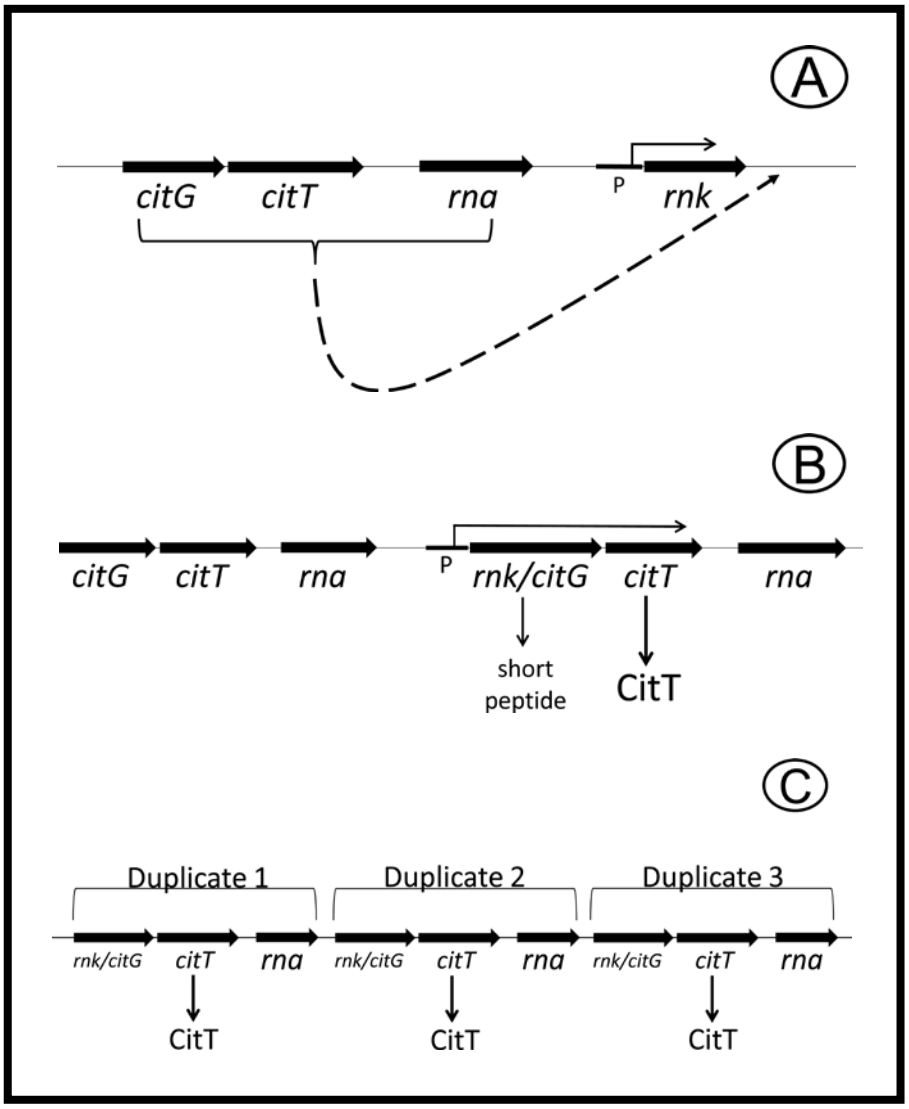


Figure 6. Illustration of the *rnk*/*citGT* fusion. (A) chromosome rearrangement moves a copy of the *citG*/*citT*/*rna* region downstream of *rnk*. (B) A fusion joins segments of *rnk* and *citG*, creating a hybrid *rnk*/*citG* region. The transcript of this region produces a short peptide with no identified function. The downstream *citT* is now under transcriptional control of the *rnk* promoter, which can function in aerobic conditions. (C) Duplication of the fusion results in multiple copies of the *rnk*/*citG* fusion and downstream *citT* and *rna*. The product of the *rna* gene is RNase I, which has not been shown to be directly relevant to the Cit⁺ phenotype.

and *dctA* mutations), some additional mutations potentially contributed to the development of the original Cit⁺ mutants. These mutations negated the effect of the earlier potentiating mutations. As mentioned above, the potentiating mutations may have established a physiological background for the Cit⁺ phenotype to develop (e.g., acetate metabolism driving the production of excess succinate for exporting), but once the *citT* and *dctA* mutations occur, the potentiating mutations are potentially detrimental.

Quandt et al. (2015) found that a group of mutations reduce or eliminate the effect of *gltA1*. Designated as *gltA2*, these secondary mutations of *gltA* reduce citrate synthase activity in the cell. This reduction of synthase activity potentially helps the mutants maintain a carbon and redox balance when catabolizing citrate. Interestingly, in the absence of citrate as a carbon and energy source, the *gltA2* mutations were generally deleterious to the bacterium (Quandt et al., 2015).

By 34,000 generations, a nonsense mutation in *aceA* also appears in most of the Cit⁺ mutants (Quandt et al., 2015). The *aceA* gene expresses isocitrate lyase, which diverts isocitrate from the Krebs cycle into the glyoxylate shunt. A nonsense mutation of *aceA* causes an inactive form of isocitrate lyase to be produced. The lack of this lyase activity enables more isocitrate to be oxidized to α-ketoglutarate (Figures 3 and 9). This mutation likely compensates for the effect of the potentiating *iclR* mutation, which increases shunting of isocitrate rather than its continued oxidation in the Krebs cycle.

Darwinian Significance

Since the initial report of the Cit⁺ mutants provides little genotypic information, this had the effect of creating a type of mystique regarding these organisms. Many assumptions and conclusions about the mutants' evolutionary significance were offered with little actual understanding of their genetics or physiology. Add to this the apparent multistage "evolutionary innovation," and the development of the Cit⁺ phenotype has become a popular "show piece" for evolutionists (Hendrickson and Rainey, 2012; Pennisi, 2013).

Among the claims are that subpopulations of *E. coli* were able to "evolve" a novel capability they did not previ-

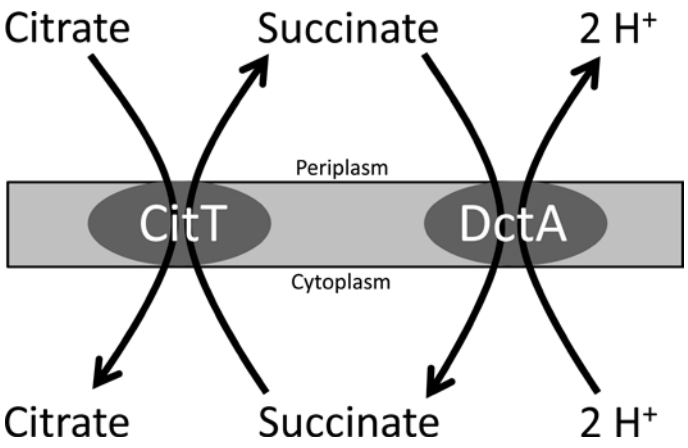


Figure 7. Coordination of CitT and DctA activity. Import of succinate by DctA provides antiporter for CitT to cotransport citrate.

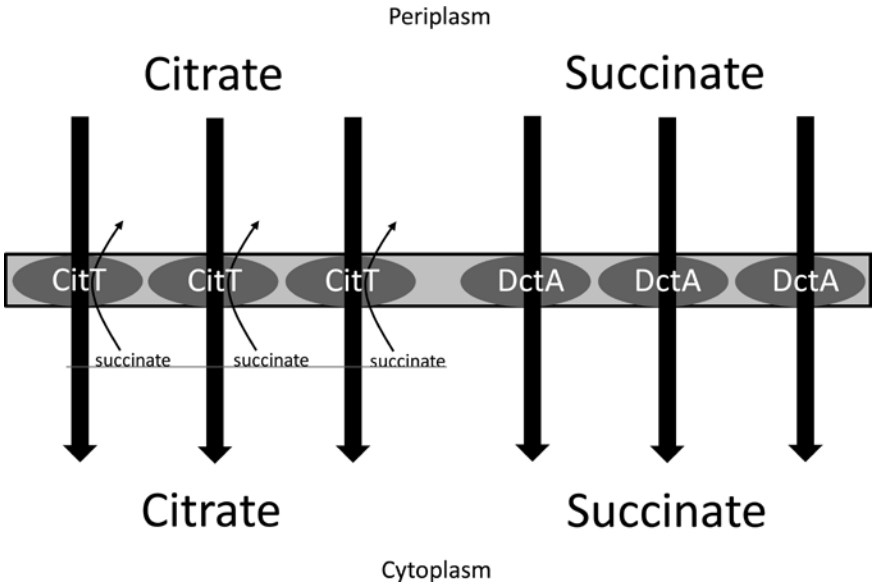


Figure 8. Epistatic relationship of CitT and DctA. High copy numbers of DctA increases transport activity of succinate into the cell, providing more succinate for antiporting by CitT. In turn, this allows high copy numbers of CitT to function, enabling higher levels of citrate to be transported into the cell. If copy numbers of DctA decline, then cellular levels of succinate decrease, which decreases CitT activity.

ously possess. It is even suggested that "a complex new function develop[ed] seemingly from scratch" (Pennisi, 2013, p. 793). Venema (2012a, 2012b) points to the results of the LTEE, and the Cit$^+$ phenotype in particular, as a "demonstration that new genes can indeed" evolve, and this is an example of a mutation that "creates a new regulatory element." Indeed, formation of these Cit$^+$ *E. coli* mutants is frequently offered as the quintessential example of evolution in action—the evolution of a new function possibly giving rise to a new species. With more genetic details of the Cit$^+$ mutants now available, this claim can be more closely assessed.

Blount et al. (2012) suggest the need for a potentiating mutation (sometime after 20,000 generations) to provide the genetic background for the ultimate development of the Cit$^+$ phenotype. Some populations of the LTEE developed a hypermutation phenotype (Barrick et al., 2009; Wielgoss et al., 2013), which presumably increases the likelihood of beneficial mutations developing and boosts the organism's fitness trajectory (Wiser et al., 2013). Interestingly, the Cit$^+$ phenotype did not arise from these hypermutable populations (Blount et al., 2008).

New Gene or Regulatory Element?

The *rnk*/*citGT* fusion captures the preexisting *rnk* promoter via chromosomal rearrangement. Gene duplication then results in multiple copies of this fusion. Rearrangements and duplications are common in enteric bacterial chromosomes (Roth et al., 1996; Matthews et al., 2011). The promoter for the *cit* operon is not part of the *citT* gene; hence, *citT* remains intact following this rearrangement. A new gene is not formed. Fusion of the 3' end of *rnk* with the 5' end of *citG* formed a hybrid *rnk*/*citG* region, which produces an 89 amino acid peptide with no reported activity (Blount et al., 2012). This does not constitute formation of a new gene, nor does the small peptide appear to have any relevance to the Cit$^+$ phenotype. Rather, functionality of both *rnk* and *citG* are lost in forming this fusion.

A new promoter is not formed either since the *rnk* promoter remains intact following the rearrangement. This preexisting *rnk* promoter is merely

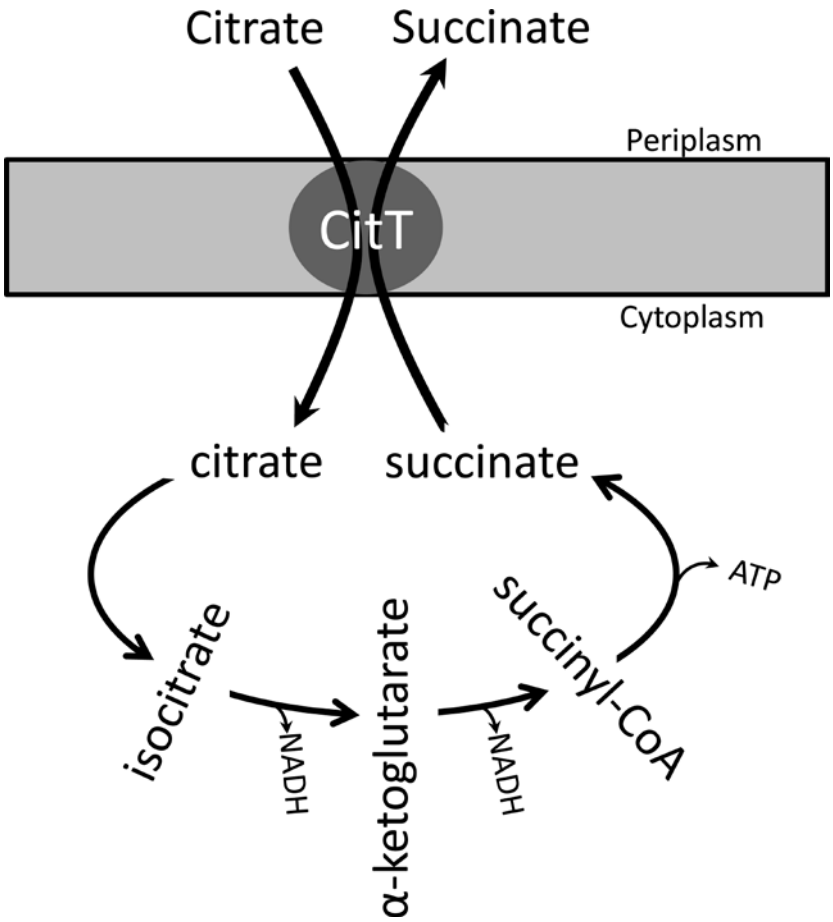


Figure 9. A possible *E. coli* metabolic pathway for the catabolism of citrate under aerobic conditions. Unlike anaerobic catabolism of citrate, it may not be necessary for *E. coli* to co-metabolize glucose. Since wild-type strains of *E. coli* do not catabolize citrate under aerobic conditions, many aspects of aerobic metabolism have yet to be investigated.

placed upstream of *citT*, where it drives transcription of both the hybrid *rnk*/*citG* region and *citT*. The fusion of *rnk*/*citG* does not form a new promoter or regulatory region.

Quandt et al. (2015) detected several mutations that apparently contribute to a favorable physiological background for subsequent development of the Cit$^+$ phenotype. Mutations of *glt*A, *iclR*, and *arcB* all can increase acetate utilization (Leiby et al., 2014), which potentially provides higher levels of C_4-dicarboxlates for cotransporting by CitT. While these mutations may constitute a necessary potentiating step, they are all loss-of-function mutations. As such, all can be categorized as antagonistic pleiotropy, but none serve as an example of "new gene" or "new regulatory element" formation.

The location of the *dct*A mutation indicates that it lies within the gene's regulatory region, altering promoter site activity. Increased levels of *dct*A transcription (as indicated by levels of mRNA) were found in the Cit$^+$ mutants compared with transcription levels in the wild-type (Quandt et al., 2014). This is consistent with loss of regulatory control of the promoter so that the gene is overexpressed.

E. coli B (the wild-type strain in the LTEE) cannot utilize C_4-dicarboxylates, such as succinate and fumarate, as energy sources (Quandt et al., 2014). This phenotype is probably attributed to a lack of *dcu*S expression in this strain of *E. coli* (Yoon et al., 2012). In a different strain (*E. coli* K-12), DcuS is part of a two-component regulatory system (DcuS-DcuR) that activates expression of C_4-dicarboxylate utilization genes, including *dct*A (Golby et al., 1999) (Figure 10A). Therefore, lack of DcuS production in *E. coli* B prevents this strain from expressing *dct*A. Quandt et al. (2014) speculate that a mutation in the Cit$^+$ cells permits expression of *dct*A, even in the absence of DcuS production (Figure 10B). The specific nature of the mutation, though, is not yet known.

Nonetheless, the *dct*A mutation apparently alters regulation of the gene. Such genotypes are common in bacteria, as promoters and other regulatory systems are easily altered by point or deletion mutations. While the exact nature of the *dct*A mutation has yet to be identified, there is little reason to assume this mutation is more than a deregulating event. Point or deletion mutations within the promoter's core site or adjacent operator regions can result in altered promoter regulation (Gourse et al. 2000; Rhodius and Mutalik, 2010). For example, increased activity is observed for some bacterial promoters following a point mutation within the -10 region (Han et al., 1991) or a single nucleotide deletion within the -35 region (Burchhardt et al., 1997).

Cells need to maintain specific promoter activity to sustain physiologically appropriate levels of certain enzymes and other proteins. Mutations within the promoter/operator region can affect the binding affinity of a regulatory protein or the RNA polymerase. Either of these mutations can potentially cause overexpression of the gene, which can upset the critical physiological balance needed by the cell. As such, these mutations

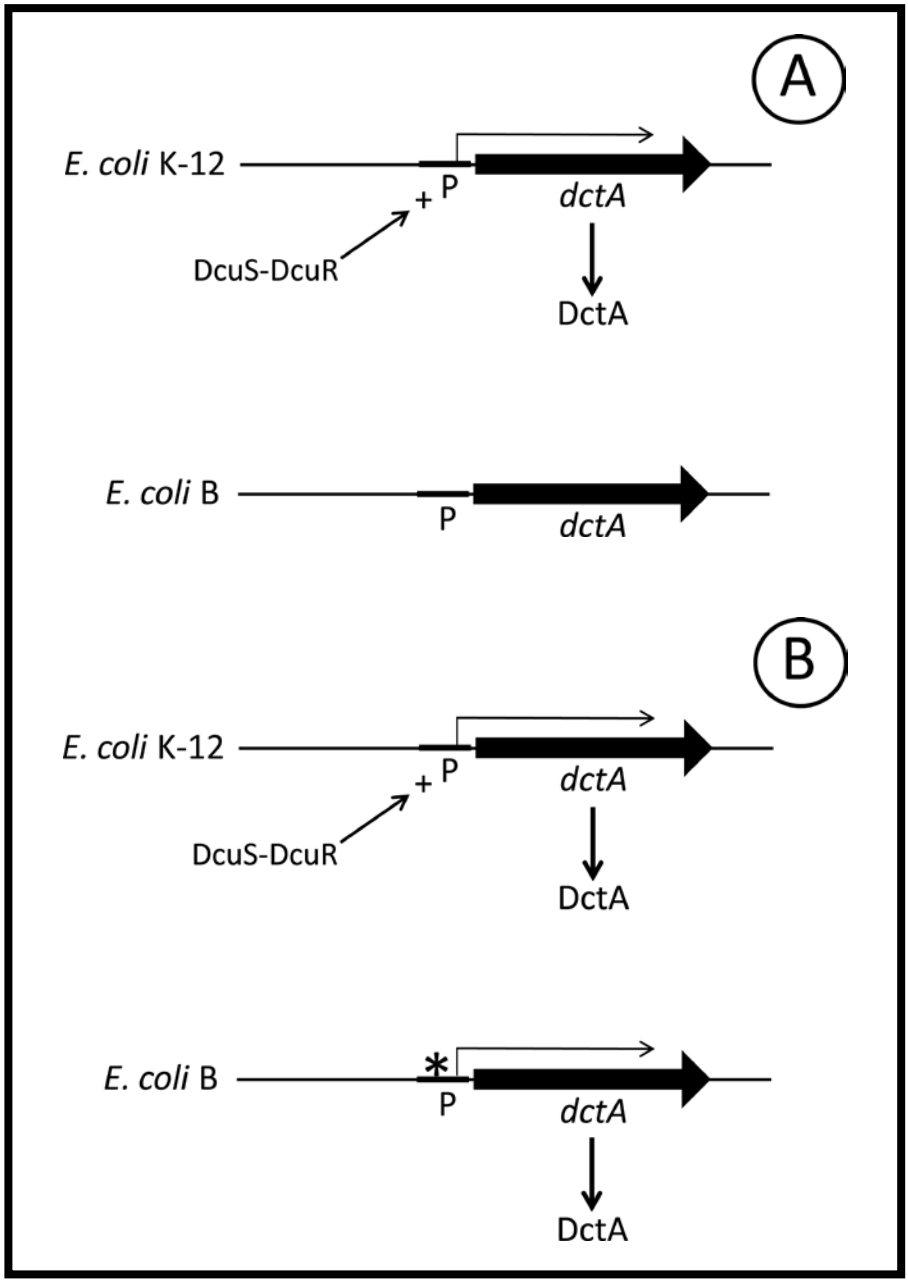


Figure 10. Comparison of *dctA* activity in *E. coli* K-12 and *E. coli* B. (A) DcuS-DcuR activates the *dctA* promoter, enabling expression of the gene in *E. coli* K-12. *E. coli* B (the LTEE wild-type strain) lacks DcuS, causing the *dctA* promoter to remain inactive. (B) In Cit$^+$ mutants, a mutation (indicated by *) in the regulatory region of the *dctA* gene enables expression of the gene even in the absence of DcuS.

can be categorized as a loss of function, since the cell can lose critical control of gene activity (e.g., overexpression is common in many cancer cells). In fact, point mutations that increase promoter activity may contribute to a variety of genetic diseases in humans (Theuns et al., 2006). This is also illustrated in the observation that the *dctA* mutation may be deleterious to the bacterium without the compensating increase of CitT production (Qunadt et al., 2014).

The subsequent secondary mutations in Cit$^+$ strains (which negate the potentiating mutations) are also loss of function. Most of the *gltA*2 mutations alter the amino acid sequence of citrate synthase, rendering the enzyme less active (Quandt et al., 2015). One of the *gltA*2 mutations (designated *gltA*2-*R*) is a point mutation near the gene's promoter region, causing a reduction in *gltA* transcription (as measured by mRNA production) (Quandt et al., 2015). In addition, the *aceA* mutation eliminates isocitrate lyase activity (Quandt et al., 2015). This compensates for the effect of the *iclR* and *arcB* mutations.

All these secondary mutations serve to compensate by reducing the activity levels of certain enzymes. Each of these secondary mutations are antagonistic pleiotrophy—beneficial to the organism as it shifts metabolism to citrate but at the expense of preexisting enzyme activity. No new genes or promoters were formed, only loss of preexisting enzyme activity and promoter function.

Therefore, the mutations detected in the Cit$^+$ organisms fail to provide a genetic mechanism for the origin of genes, promoter or operator sites, or the origin of any regulatory elements. Instead, they serve as an example of how loss of preexisting activity or the overexpression of specific genes can provide an adaptive advantage in specialized conditions.

New Function?

From their analysis, Blount et al. (2012, p. 517) conclude that the Cit$^+$ phenotype results from "the multi-step origin of a key innovation" producing "new functions." Yet, *E. coli* already possesses the ability to utilize citrate for energy. The genes for all the needed enzyme and transport systems are already present.

Potentially, the most straightforward mutation to achieve aerobic expression of the *cit* operon would be one that directly alters its regulation, thereby enabling

aerobic expression of the operon. The absence of this genotype in any Cit$^+$ mutants suggests such alterations may have overriding negative consequences. For example, the aerobic/anaerobic regulation of the *cit* operon is apparently a function of the redox sensitivity of CitA. In high redox conditions (i.e., aerobic), CitA is inactive. Perhaps eliminating this redox sensitivity reduces the overall activity of CitA, thus the *cit* operon would still remain inactive in aerobic conditions.

On the other hand, any mutation that places *citT* under control of an aerobically functioning promoter can offer at least a low-level Cit$^+$ phenotype. What likely makes the *rnk*/*citGT* fusion the most effective of the various promoter captures is that this fusion can occur in tandem repeats. Apparently the increased number of CitT copies that result from this duplication is necessary for a strong Cit$^+$ phenotype. The *rnk*/*citGT* fusion moves *citT* upstream of the *rnk* promoter, but the promoter and the gene remain unchanged.

The *dctA* mutation apparently alters the promoter activity of the gene. This alteration removes some type of regulatory control of the gene, causing overexpression. As discussed above, this is a loss-of-function mutation.

The benefit of this *dctA* mutation is related to its epistatic relationship with the *rnk*/*citGT* fusion. Epistatic interactions of genes and mutations are not unique, per se, as bacteria have many—some with a positive affect (e.g., Trindade et al., 2009) and some a negative affect (e.g., Khan et al., 2011). Interestingly, negative epistasis appears to be more common in eukaryotic organisms than in bacteria (Xu et al., 2012). Hence, bacteria may be more capable of benefiting from an epistatic interaction of mutations than other organisms.

While a number of different mutations were found in many of the Cit$^+$ mutants, only the *dctA* mutation and *rnk*/*citGT* fusion were necessary for the full Cit$^+$ phenotype (Quandt et al., 2014). Yet, neither of these mutations can appropriately be classified as "birth" of a "complex new function … from scratch." A preexisting gene "captures" a preexisting promoter, and regulatory control of another preexisting promoter is lost.

Ultimately, the development of the Cit$^+$ phenotype results from the reorganization of preexisting genes, loss of preexisting enzyme activity, and elimination of preexisting regulatory control. From a research perspective, the Cit$^+$ mutants offer a very interesting study in bacterial adaptation and genetic versatility (and certainly a "yeoman task" to elucidate). However, if the wild-type *E. coli* strain did not already possess these genetic systems (such as *citT* and *dctA*), there have been no mutations observed during the entire LTEE that would have generated the Cit$^+$ phenotype "from scratch."

Since both glucose and citrate are present in the medium at very low levels, it is not surprising that any mutant that can effectively use both of these as an energy and carbon source would have a competitive advantage against the wild-type strain. Yet, the Cit$^+$ phenotype is possibly less advantageous to the bacterium than initially concluded. In the original study, the Cit$^-$ population eventually became extinct in the presence of the Cit$^+$ strains. This was assumed to be a consequence of competition from the more-fit Cit$^+$ population but may have actually been a result of random experimental factors. In subsequent studies, populations of Cit$^-$ cells (adapted to the LTEE medium) coexisted with Cit$^+$ mutants, and no extinction occurred even after 2,500 generations (Turner et al., 2015). While the Cit$^+$ mutants would appear to have a distinct growth advantage over the Cit$^-$ population, the advantage is apparently less than would be predicted. Perhaps the heavy mutational load of the Cit$^+$ strains exerts a sufficient toll that reduces the overall physiological benefit of being able to utilize both glucose and citrate. Perhaps other factors will eventually be elucidated.

Bacteria as a Model

In the game of adaptation, bacteria have a decided advantage over most other organisms. For example, bacteria can sustain and survive (at least temporarily) a higher mutation rate than vertebrates (Denamur and Matic, 2006; Linz et al., 2014). In fact, the hypermutator phenotype that developed during the LTEE would potentially be lethal to vertebrate populations (Sniegowski et al., 2000). For bacteria, however, this high mutation rate may increase the probability that an adaptive mutation will occur (Wielgoss et al., 2013).

Bacteria also have a much faster generation rate than vertebrates (optimal generation time frequently measured in minutes rather than years or decades). In fact, they outnumber all sexually reproducing populations by multiple orders of magnitude (Crawford, 2007), enabling them to roll the "mutation dice" quadrillions of times each day. In addition, bacteria have several genetic "tricks" they can employ for adaptation, such as a high level of genome plasticity and a vast pool of horizontally transferrable "plug-and-play" genes (Blount, 2015; Lenski, 2004). Combined with a potential hypermutation phenotype and haploid genetics (i.e., one chromosome, no heterozygous genotypes), these characteristics enable bacteria to often quickly adapt to different, even hostile, environments. As such, they are excellent engines of adaptation, explaining why they can be found in virtually every part of earth's biosphere. All these characteristics also enable bacteria to pay a much higher "cost of substitution" than most organisms (for a more detailed discussion of this principle see ReMine, 2005, 2006).

This raises questions of how applicable bacterial adaptation studies are to "higher" organisms, such as plants or animals. As haploid asexual organisms, their reproductive biology and inheritance differ significantly from sexual reproduction and inheritance. Thus, the ability

of bacteria to genetically accomplish an adaptation does not mean other organisms can repeat this achievement.

For example, attempting to apply the LTEE design to diploid sexual organisms, Burke et al. (2010) conducted a long-term experiment with *Drosophila*. They applied a fairly strong "selection" protocol for accelerated development and early fertility of the flies. After over 600 generations of "selection," the researchers found several phenotypes present in the experimental population, including a 20% faster rate of development from egg to adult. Yet, they found no unconditional beneficial mutation had become fixed within the population. Since the experimental conditions are much more intense than what *Drosophila* would experience in a natural setting, the researchers note that it is even less likely any beneficial mutation would become fixed within a wild population in the same time frame. A subsequent experiment of *Drosophila* also failed to detect significant fixation of beneficial mutations following numerous generations of "selection" (Orozco-terWengel et al., 2012). This illustrates how difficult it is for *Drosophila* to generate and fix just a single beneficial mutation, even under significant "selective" pressure.

Conclusions

From a microbiology and genetics perspective, the Cit$^+$ mutants are a fascinating collection of organisms. They offer many informative insights into the preprogrammed ability of bacteria to adapt to environmental limitations (such as the limited food source in the LTEE base medium). The multiple steps involved in the adaptation makes the mutants attractive as a mechanism for Darwinian evolution. However, as intriguing as this multistep sequence is as a study model, careful analysis of the genotype of these mutants reveals they are not an appropriate example of a "*de novo* origination of genes" or an organism building a new function "from scratch."

What new gene was formed? Genes involved in the potentiating step (*gltA*, *iclR*, and *arcB*) and the secondary mutations of the refining step (*gltA* and *aceA*) are already present in the wild-type bacterium. Mutations of these genes reduce gene expression, enzyme activity, and regulatory controls. All these mutations would be categorized as degenerative—the opposite of "*de nova* origination of genes." What is more, all genes involved in providing the full Cit$^+$ phenotype (*rnk*, *citG*, *citT*, and *dctA*) were also already present in the wild-type organism. Moving *citT* upstream of a different promoter does not constitute formation of a new gene. In addition, alteration of the regulatory site upstream of *dctA* does not constitute a new gene, only the elimination of the transcriptional control. Both *citT* and *dctA* remain structurally intact and their transcripts unchanged. Also, the *rnk*/*citG* fusion actually eliminates *rnk* and *citG* as functioning genes.

Where is the new regulatory element? The *gltA2-R* mutation reduces activity of the *gltA* promoter. This mutation lowers transcription levels of the gene. The reduction of transcription activity of a preexisting promoter does not provide a genetic mechanism for the origin of new promoters. Also, the mutations of *iclR*, and *arcB* eliminate the function of regulatory elements rather than form new elements. Note as well, the different promoters (those from IS3, *fimB*, and *rnk*) captured for aerobic expression of *citT* were already present in the wild-type strain. Each of the promoters remains intact and unchanged. The "capture" merely places *citT* adjacent to these preexisting promoters. In the same context, a mutation within the *dctA* promoter/operator region did not form a new promoter; rather the mutation eliminates preexisting regulatory control of that promoter.

What is the new function? Utilization of citrate is not a new function for *E. coli*, only the environmental condition of that utilization has changed (i.e., aerobic vs. anaerobic). This change occurred by (1) eliminating the regulation of a preexisting system (i.e., deregulating *dctA*) and (2) removing *citT* from a preexisting regulatory control and placing it under a different preexisting control. Therefore, claims of a new function are, at best, context dependent.

Whether argued as a new function or not, fusion of a preexisting gene to a preexisting promoter fails to offer a genetic mechanism for the origin of either the gene or the promoter. Nor does elimination of a preexisting regulatory system serve as an example of how regulatory systems originated. Furthermore, duplication of *citT* merely copies a preexisting gene and hence provides no insight into the origin of the duplicated gene. Even if Darwinists claim gene duplication allows for subsequent "evolution" of one of the duplicated genes, this did not occur in the Cit$^+$ mutants. The duplicated *citT* remains the same gene.

The genotypes of the Cit$^+$ strains are interesting examples of how bacteria can rearrange preexisting genes, reduce preexisting enzyme activity, and alter regulation of preexisting promoters. However, mutations within these strains did not form new genes or promoters "from scratch." Rather, they serve as another example of antagonistic pleiotropy, or loss of preexisting systems providing a temporal benefit. As such, the mutants fail to provide a genetic example of how genes, promoters, and regulatory systems originated. Thus, the Cit$^+$ mutants are not an example of the required genetic mechanism for neo-Darwinian transformation enabling universal common descent.

Despite being an excellent engine for adaptation, bacteria fail to provide any examples of a genetic mechanism for new genes and regulatory controls. If bacteria are unable to offer such a mechanism, what basis is there to presume that it can be attained by lesser

engines, such as mice, trees, fruit flies, and primates? The literature is filled with examples of phenotypic changes but barren of genetic examples enabling universal common descent. The failure of the Cit$^+$ mutants to provide such an example further reinforces the general failure of Darwinian evolution as an explanatory tool for the origin of genetic systems.

In contrast, the Cit$^+$ mutants fit appropriately within the predictions of a creation model. Organisms, such as bacteria, are preprogrammed to adapt to differing environments. This adaptation can involve eliminating preexisting regulatory controls, chromosome rearrangement, and even using external elements (such as IS3) for altered gene activity. Yet, the alterations were limited in scope and did not result in the *de nova* origination of new genes or regulatory elements.

References

Allers, K., G. Hütter, J. Hofmann, C. Loddenkemper, K. Rieger, E. Thiel, and T. Schneider. 2011. Evidence for the cure of HIV infection by CCR5Δ32/Δ32 stem cell transplantation. *Blood* 117:2791–2799.

Anderson, K. 2005. Is bacterial resistance to antibiotics an appropriate example of evolutionary change? *Creation Research Society Quarterly* 41:318–326.

Anderson, K., and G. Purdom. 2008. A creationist perspective of beneficial mutations in bacteria. In Snelling, A.A. (editor), *Proceedings of the 5th International Conference on Creationism*, pp. 73–86. Creation Science Fellowship, Pittsburgh, PA, and Institute for Creation Research, Dallas, TX.

Barrick, J.E., D.S. Yu, S.H. Yoon, H. Jeong, T.K. Oh, D. Schneider, R.E. Lenski, and J.F. Kim. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461:1243–1247.

Blount, Z.D. 2015. The unexhausted potential of *E. coli*. *eLife* 4:e05826 (doi:10.7554/eLife.05826).

Blount, Z.D., C.Z. Borland, and R.E. Lenski. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA* 105:7899–7906.

Blount, Z.D., J.E. Barrick, C.J. Davidson, and R.E. Lenski. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* 489:513–518.

Burchhardt, G., I. Schmidt, H. Cuypers, L. Petruschka, A. Völker, and H. Herrmann. 1997. Studies on spontaneous promoter-up mutations in the transcriptional activator-encoding gene *phlR* and their effects on the degradation of phenol in *Escherichia coli* and *Pseudomonas putida*. *Molecular and General Genetics* 254:539–547.

Burke, M.K., J.P. Dunham, P. Shahrestani, K.R. Thornton, M.R. Rose, and A.D. Long. 2010. Genome-wide analysis of long-term evolution experiment with *Drosophila*. *Nature* 467:587–590.

Carlin, J.L. 2011. Mutations are the raw materials of evolution. *Nature Education Knowledge* 3(10): 10.

Charlier, D., J. Piette, and N. Glansdorff. 1982. IS3 can function as a mobile promoter in *E. coli*. *Nucleic Acids Research* 10:5935–5948.

Cholera, R., N.J. Brittain, M.R. Gillrie, T.M. Lopera-Mesa, S.A.S. Diakité, T. Arie, M.A. Krause, A. Guindo, A. Tubman, H. Fujioka, D.A. Diallo, O.K. Doumbo, M. Ho, T.E. Wellems, R.M. Fairhurst. 2008. Impaired cytoadherence of *Plasmodium falciparum*-infected erythrocytes containing sickle hemoglobin. *Proceedings of the National Academy of Sciences, USA* 105:991–996.

Cooper, T.F., D.E. Rozen, and R.E. Lenski. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA* 100:1072–1077.

Cooper, V.S., and R.E. Lenski. 2000. The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* 407:736–739.

Cooper, V.S., A.F. Bennett, and R.E. Lenski. 2001a. Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. *Evolution* 55:889–896.

Cooper, V.S., D. Schneider, M. Blot, and R.E. Lenski. 2001b. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *Journal of Bacteriology* 183:2834–2841.

Coyne, J.A. 2009. *Why Evolution Is True*. Penguin Books, New York, NY.

Crawford, D.H. 2007. *Deadly Companions: How Microbes Shaped Our History*. Oxford University Press, Oxford, UK.

Cronan, J., and D. Laporte. 2006. Tricarboxylic acid cycle and glyoxylate bypass. *EcoSal Plus* doi:10.1128/ecosalplus.3.5.2.

Davies, S.J., P. Golby, D. Omrani, S.A. Broad, V.L. Harrington, J.R. Guest, D.J. Kelly, and S.C. Andrews. 1999. Inactivation and regulation of the aerobic C_4-dicarboxylate transport (*dctA*) gene of *Escherichia coli*. *Journal of Bacteriology* 181:5624–5635.

Dawkins, R. 1996. *Climbing Mount Improbable*. W.W. Norton & Company, New York, NY.

Denamur, E., and I. Matic. 2006. Evolution of mutation rates in bacteria. *Molecular Microbiology* 60:820–827.

Depew, D.J., and B.H. Weber. 2011. The fate of Darwinism: evolution after the modern synthesis. *Biological Theory* 6:89–102.

Fry, J.D. 1993. The "general vigor" problem: can antagonistic pleiotropy be detected when genetic covariances are positive? *Evolution* 47:327–333.

Golby, P., S. Davies, D.J. Kelly, J.R. Guest, and S.C. Andrews. 1999. Identification and characterization of a two-component sensor-kinase and response-regulator system (Dcu-S-DcuR) controlling gene expression in response to C_4-dicarboxylates in *Escherichia coli*.

Journal of Bacteriology 181:1238–1248.

Gourse, R.L., W. Ross, and T. Gaal. 2000. UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Molecular Microbiology* 37:687–695.

Han, C.Y., I.P. Crawford, and C.S. Harwood. 1991. Up-promoter mutations in the *trpBA* operon of *Pseudomonas aeruginosa*. *Journal of Bacteriology* 173:3756–3762.

Hendrickson, H., and P.B. Rainey. 2012. How the unicorn got its horn. *Nature* 489:504–505.

Khan, A.I., D.M. Dinh, D. Schneider, R.E. Lenski, and T.F. Cooper. 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332:1193–1196.

Leiby, N., C.J. Marx, and N.A. Moran. 2014. Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in Escherichia coli. *PLoS Biology* 12:e1001789.

Lengeler, J.W., G. Drews, and H.G. Schlegel. 1999. *Biology of the Prokaryotes*. Blackwell Science, New York, NY.

Lenski, R. 2003. Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. In Janick, J. (editor), *Plant Breeding Reviews: Long-term Selection: Crops, Animals, and Bacteria*. Vol. 2, Part 2, pp. 225–265. Wiley Press, New York, NY.

Lenski, R. 2004. The future of evolutionary biology. *Ludus Vitalis* 12:67–68.

Lenski, R. 2010. Experimental evolution. http://myxo.css.msu.edu/ecoli/dm25liquid.html (accessed October 9, 2015).

Lenski, R.E., and M. Travisano. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences, USA*. 91:6808–6814.

Linz, B., H.M. Windsor, J.J. McGraw, L.M. Hansen, J.P. Gajewski, L.P. Tomsho, C.M. Hake, J.V. Solnick, S.C. Schuster, and B.J. Marshall. 2014. A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques. *Nature Communication* 5:4165.

Margulis, L., and D. Sagan. 2002. *Acquiring Genomes: A Theory of the Origin of Species*. Basic Books, New York, NY.

Matthews, T.D., W. Rabsch, and S. Maloy. 2011. Chromosomal rearrangements in *Salmonella enterica* serovar Typhi strains isolated from asymptomatic human carriers. *mBio*2(3): e00060–11. doi:10.1128/mBio.00060–11.

Merlin, F. 2010. Evolutionary chance mutation: a defense of the modern synthesis' consensus view. *Philosophy and Theory in Biology* 2:e103.

Noble, D. 2015. Evolution beyond neo-Darwinism: a new conceptual framework. *Journal of Experimental Biology* 218:7–13.

Novak, M., T. Pfeiffer, R.E. Lenksi, U. Sauer, and S. Bonhoeffer. 2006. Experimental tests for an evolutionary trade-off between growth rate and yield in *E. coli*. *The American Naturalist* 168:242–251.

Olson, M.V. 1999. When less is more: gene loss as an engine of evolutionary change. *American Journal of Human Genetics* 64:18–23.

Orozco-terWengel, P., M. Kapun, V. Nolte, R. Kofler, T. Flatt, and C. Schloetterer. 2012. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Molecular Ecology* 21:4931–4941.

Ostrowski, E.A., D.E. Rozen, and R.E. Lenski. 2005. Pleiotrophic effects of beneficial mutations in *Escherichia coli*. 59:2343–2352.

Patterson, C. 1978. *Evolution*. Cornell University Press, Ithaca, NY.

Pelosi, L., L. Kühn, D. Guetta, J. Garin, J. Geiselmann, R.E. Lenski, and D. Schneider. 2006. Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. *Genetics* 173:1851–1869.

Pennisi, E. 2013. The man who bottled evolution. *Science* 342:790–793.

Philippe, N., L. Pelosi, R.E. Lenski, and D. Schneider. 2009. Evolution of penicillin-binding protein 2 concentration and cell shape during a long-term experiment with *Escherichia coli*. 191:909–921.

Pos, K.M., P. Dimroth, and M. Bott. 1998. The *Escherichia coli* citrate carrier CitT: a member of a novel eubacterial transporter family related to the 2-oxoglutarate/malate translocator from spinach chloroplasts. *Journal of Bacteriology* 180:4160–4165.

Quandt, E.M., D.E. Deatherage, A.D. Ellington, G. Georgiou, and J. Barrick. 2014. Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA* 111:2217–2222.

Quandt, E.M., J. Gollihar, Z.D. Blount, A.D. Ellington, G. Georgiou, and J.E. Barrick. 2015. Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. *eLife* 4:e09696. doi:10.7554/eLife.09696

ReMine, W. 2005. Cost theory and the cost of substitution—a clarification. *Journal of Creation* 19:113–125.

ReMine, W. 2006. More precise calculations of the cost of substitution. *Creation Research Society Quarterly* 43:111–120.

Rhodius, V.A., and V.K. Mutalik. 2010. Predicting the strength and function for promoters of the *Escherichia coli* alternative sigma factor, σ^E. *Proceedings of the National Academy of Sciences, USA* 107:2854–2859.

Roff, D.A., and D.J. Fairborn. 2006. The evolution of trade-offs: where are we? *Journal of Evolutionary Biology* 20:433–447.

Rose, M.R. 1985. Life history evolution with antagonistic pleiotrophy and overlapping generations. *Theoretical Population Biology* 28:342–358.

Roth, J.R., N. Benson, T. Galitski, K. Haack, J.G. Lawrence, and L. Miesel. 1996. Rearrangements of the bacteria chromosome: formation and applications. In Neidhardt, F.C. (editor), *Escherichia coli and Salmonella: Cellular and Molecular*

Biology, vol. 2, pp. 2256–2276. ASM Press, Washington DC.

Ruiz-Orera, J., J. Hernandez-Rodriquez, C. Chiva, E. Sabidó, I. Kondova, R. Bontrop, T. Marqués-Bonet, and M. Mar Albá. 2015. Origins of *de novo* genes in human and chimpanzee. *PLoS Genetics* 11(12): e1005721.

Scheu, P.D., J. Witan, M. Rauschmeier, S. Graf, Y.-F. Liao, A. Ebert-Jung, T. Basché, W. Erker, and G. Unden. 2012. CitA/CitB two-component system regulating citrate fermentation in

Escherichia coli and its relation to the DcuS/DcuR system *in vivo*. *Journal of Bacteriology* 194:636–645.

Scheutz, F., and N.A. Strockbine. 2005. Genus I. *Escherichia*, Castellani and Chalmers 1919. In Garrity, G.M., D.J. Brenner, N.R. Kreig, and J.R. Staley (editors), *Bergey's Manual of Systematic Bacteriology*, *Volume Two: The Proteobacteria*, pp 607–624. Springer, New York, NY.

Sleight, S.C., N.S. Wigginton, and R.E. Lenski. 2006. Increased susceptibility to repeated freeze-thaw cycles in *Escherichia coli* following long-term evolution in a benign environment. *BMC Evolutionary Biology* 6:104 (doi:10.1186/1471–2148–6-104).

Sniegowski, P.D., P.J. Gerrish, T. Johnson, and A. Shaver. 2000. The evolution of mutation rates: separating causes from consequences. *BioEssays* 22:1057–1066.

Steinmetz, P.A., S. Wörner, and G. Unden. 2014. Differentiation of DctA and DcuS function in the DctA/DcuS sensor complex of *Escherichia coli*: function of DctA as an activity switch and of DcuS as the C_4-dicarboxylate sensor. *Molecular Microbiology* 94:218–229.

Theuns, J., N. Brouwers, S. Engelborghs, K. Sleegers, V. Bogaerts, E. Corsmit, T. De Pooter, C.M. van Duijn, P.P. De Deyn, and C. Van Broeckhoven. 2006. Promoter mutations that increase amyloid precursor-protein expression are associated with Alzheimer disease. *The American Journal of Human Genetics* 78:936–946.

Trindade, S., A. Sousa, K.B. Xavier, F. Dionisio, M.G. Ferreira, and I. Gordo. 2009. Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genetics* 7(7): e1000578.

Turner, C.B., Z.D. Blount, and R.E. Lenski. 2015. Replaying evolution to test the cause of extinction of one ecotype in an experimentally evolved population. *PLoS One* 10(11): e0142050.

Venema, D. 2012a. Behe, Lenski and the "edge" of evolution, part 2: gaining a new function. http://biologos.org/blogs/dennis-venema-letters-to-the-duchess/behe-lenski-and-the-edge-of-evolution-part-2-gaining-a-new-function (accessed December 18, 2015).

Venema, D. 2012b. Behe, Lenski and the "edge" of evolution, part 5: mixing and matching. http://biologos.org/blogs/dennis-venema-letters-to-the-duchess/behe-lenski-and-the-edge-of-evolution-part-5-mixing-and-matching (accessed December 17, 2015).

Wielgoss, S., J.E. Barrick, O. Tenaillone, M.J. Wiser, W.J. Dittmarh, S. Cruveiller, B. Chane-Woon-Ming, C. Médigue, R.E. Lenski, and D. Schneider. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proceedings of the National Academy of Sciences, USA* 110:222–227.

Wiser, M.J., N. Ribeck, and R.E. Lenski. 2013. Long-term dynamics of adaptation in asexual populations. *Science* 342:1364–1367.

Wolfe, A.J. 2005. The acetate switch. *Microbiology and Molecular Biology Reviews* 69:12–50.

Xu, L., B. Barker, and Z. Gu. 2012. Dynamic epistasis for different alleles of the same gene. *Proceedings of the National Academy of Sciences, USA* 109:10420–10425.

Yamamoto, K., F. Matsumoto, T. Oshima, N. Fujita, N. Ogasawara, and A. Ishihama. 2008. Anaerobic regulation of citrate fermentation by CitAB in *Escherichia coli*. *Bioscience, Biotechnology, and Biochemistry* 72:3011–3014.

Yamamota, K., F. Matsumoto, S. Minagawa, T. Oshima, N. Fujita, N. Ogasawara, and A. Ishihama. 2009. Characterization of CitA-CitB signal transduction activating genes involved in anaerobic citrate catabolism in *Escherichia coli*. *Bioscience, Biotechnology, and Biochemistry* 73:346–350.

Yoon, S.H., M.-J. Han, H. Jeong, C.H. Lee, X.-X. Xia, D.-H. Lee, J.H. Shim, S.Y. Lee, T.K. Oh, and J.F. Kim. 2012. Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. *Genome Biology* 13(5): R37.

Zimmer, C. 2001. *Evolution: The Triumph of an Idea.* HarperCollins, New York, NY.

Addendum

A recent study demonstrates that *E. coli* can acquire the needed *citT* and *dctA* mutations in less than 100 generations (Van Hofwegen et al., 2016). Employing a different selection protocol than the LTEE, these researchers repeatedly obtained Cit$^+$ mutants. They conclude that the rarity of the Cit$^+$ mutants during the LTEE was strictly an artifact of the experimental conditions. Interestingly, Van Hofwegen et al. (2016) also arrive at the conclusion that no new genes were formed in the generation of the Cit$^+$ mutants. In fact, *E. coli* adaptation generally involves altering the regulatory control of pre-existing genes (Zinser and Kolter, 2004), and formation of new genes remains largely undocumented.

References

Van Hofwegen, D.J., C.J. Hovde, and S.A. Minnich. 2016. Rapid evolution of citrate utilization by *Escherichia coli* by direct selection requires *citT* and *dctA*. *Journal of Bacteriology* 198:1022–1034.

Zinser, E.R., and R. Kolter. 2004. *Escherichia coli* evolution during stationary phase. *Research Microbiology* 155:328–336.

Media Reviews

The Making of the Fittest: Evolving Switches, Evolving Bodies (Video)

Howard Hughes Medical Institute Biointeractive,

http://www.hhmi.org/biointeractive/making-fittest-evolving-switches-evolving-bodies

It has often been said that natural selection may explain the survival of the fittest, but it does not explain the arrival of the fittest. This video by the Howard Hughes Medical Institute (HHMI) attempts to address the latter. HHMI hosts a number of videos that promote both science and evolutionary philosophy, including this video that takes a look at the genetics underlying the adaptive loss of the pelvic spines in freshwater threespine sticklebacks.

The video is well done. There is beautiful natural scenery, interviews conducted by Sean Carroll with two researchers, and quality graphics that are woven together in an engaging way. Carroll is a molecular biologist who has conducted research on gene regulation in fruit fly development. In 2010 he was named vice president for science education of the Howard Hughes Medical Center, and he has played an important role in the production of the HHMI Biointeractive videos.

The first scientist interviewed is Mike Bell, who has done important work evaluating patterns of morphological variation in threespine sticklebacks. It is believed that these freshwater sticklebacks are descended from anadromous sticklebacks, which migrate from the ocean to freshwater streams to spawn. Somehow, they were cut off from the ocean and have adapted to full-time life in their current freshwater habitat. In doing so they lost their pelvic spines, which, though helpful in the ocean, can be a liability in the lake as hungry dragonfly larva grab them when attempting to make a meal of the fish. Except for the timescale, this proposed history is compatible with a biblical worldview.

Next, David Kingsley is interviewed, and the fascinating story is told of how he identified the underlying genetic difference that accounts for the loss of pelvic spines. Interestingly, the threespine stickleback is designed with multiple regulatory elements that control expression of the identified gene (*pitx1*) in different tissues. One regulatory region was deleted in the freshwater fish, and further experiments confirmed that it controlled expression of the gene in the pelvic region. The fact that other regulatory regions control *pitx1* expression in other critical tissues allows for the loss of this trait without killing the fish. What incredible design that allows for adaptation! Yet those involved in the video seem to completely miss this evidence of design because of their evolutionary worldview.

In the case of sticklebacks, freshwater species from a number of locations around the world have the same regulatory region deleted. Thus, it appears that the same basic adaptive strategy has been used multiple times. While at one time evolutionists believed this type of repeated evolution unthinkable because of their underlying assumptions of random mutation and natural selection, they have now been forced to recognize that this does occur (Brodie, 2010). Rather than seriously scrutinize their assumptions, biologists have generally marveled at how this shows the power of natural selection. In doing so, they ignore the realistic mathematical modeling that demonstrates that natural selection does not have these magical powers; it is not efficient at fixing most beneficial mutations or removing most deleterious ones (Lightner, 2015). While being enamored by the supposed powers of natural selection, they are distracted from productive research questions that could enable them to find the underlying basis of this remarkably repeatable genetic transformation made by these fish when trapped in a freshwater environment.

The video story line turns back to Mike Bell and the work he has done looking for stickleback fossils in Nevada. There are tens of thousands of layers of rock in a dried-up lake bed that contains many stickleback fossils. The deeper layers have an abundance of sticklebacks with spines, but there is a sudden shift to fish without spines after a few thousand layers. Since the layers are interpreted as being annual (varves), it is claimed this represents thousands of years. Yet work

by Mike Bell and others suggests these transformations frequently occur within a few decades (Bell et al., 2004; Lescak et al., 2015).

Notice that the observed transformation documented in the literature is happening several orders of magnitude faster than what is claimed by the evolutionists based on the fossil record. Something is wrong, namely the assumption that the layers are annual (varves). There are other dried up lake beds (e.g., Green River Formation in Wyoming) where it is very clear the layers cannot be annual (Oard and Whitmore, 2006). Also, the excellent preservation makes it clear that the fish were deposited rapidly, before significant decay or scavenging could occur (Whitmore, 2006). So empirical evidence gives us strong reason to doubt the age assigned by the evolutionists; in reality it occurred well within a biblical timeframe.

At the beginning of the video Sean Carroll mentions that the changes in sticklebacks provide evidence of how all creatures evolve. If this is so, it is powerful evidence against molecules-to-man evolution. For these sticklebacks, traits (e.g., body armor and pelvic spines) are repeatedly lost when the fish become trapped in a freshwater environment and, in the case of pelvic spines, the underlying regulatory genetic sequence is deleted. These are changes headed in the wrong direction to support evolutionary beliefs on origins. Changes in regulatory sequences require that such sequences exist and that they are designed in such a way that adaptive change is possible. Thus, logically, life must have begun with an astonishingly complex design that allows for adaptive changes. This is consistent with biblical creation, not evolution. Further, the changes occur much more rapidly and predictably than evolutionists had previously thought possible. This suggests that designed mechanisms, not just random mutation and natural selection, are likely involved.

References

Bell, M.A., W.E. Aguirre, and N.J. Buck. 2004. Twelve years of contemporary armor evolution in a threespine stickleback population. *Evolution* 58(4): 814–824.

Brodie, E.D. 2010. Convergent evolution: pick your poison carefully. *Current Biology* 20(4): R152–154.

Lescak, E.A., S.L. Bassham, J. Catchen, O. Gelmond, M.L. Sherbick, F.A. von Hippel, and W.A. Cresko. 2015. Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences* 112(52): E7204–7212.

Lightner, J.K. 2015. Matters of fact … a creationist view of natural selection. *Creation Matters* 20(5): 6–7.

Oard, M.J., and J. H. Whitmore. 2016. The Green River Formation of the west-central United States: Flood or post-Flood? *Journal of Creation* 20(1): 46–49; see comment in reference 15.

Whitmore, J. 2006. Exploding fish: evidence for rapid burial. *Answers Magazine* 1(2): 27–31.

Jean Lightner
jklightner@gmail.com

The Making of the Fittest: Natural Selection and Adaptation (Video)

Howard Hughes Medical Institute Biointeractive,

http://www.hhmi.org/biointeractive/making-fittest-natural-selection-and-adaptation

This video is one of the many Biointeractive videos that promote both science and an evolutionary worldview. Sean Carroll accompanies Michael Nachman as they discuss the adaptive coloration patterns in pocket mice in New Mexico. Carroll strongly promotes evolution, and in particular natural selection, as the major mechanism by which adaptive evolutionary changes take place. He feels these mice are an excellent example of adaptation by natural selection.

Nachman has studied mice on the black rocks of a lava flow known to be about 1000 years old, as well as mice found in the surrounding desert. The mice on the lava flow are black and blend in well with the color of the rocks. The mice living on the lighter rocks of the desert have a lighter color. In each case, the matching color makes them less of a target for predators.

Genetic evaluation has shown that black mice have a mutation in the *mc1r* gene. Four nucleotide bases differ between them and their light-colored relatives. This is an example of rapid adaptation. Interestingly, black mice from older lava flows elsewhere in the Southwest do not have this mutation. Thus, while we see rapid adaptation, much like in the sticklebacks in the *Evolving Switches, Evolving Bodies* video, the gene targeted is not always the same.

Again, the evolutionists in this video do not seem to recognize that this adaptation required a preexisting complex system that was designed in a way to allow for adaptive color changes. Most mice can produce both a lighter pigment (yellow to red) and a darker pigment (dark brown to black). In fact, if you look at a hair from the back of a light-colored mouse, you will notice a wide light band between two dark ones. The light band is missing from the dark mice, since they produce only the dark pigment (Nachman, 2005).

The mutation basically keeps the signal for the darker pigment turned on, making the mc1r unresponsive to the upstream signal that would cause the banding pattern. So while it is adaptive, it takes advantage of preexisting complexity. This is consistent with the biblical worldview of an all-wise Creator who made and sustains life (even in this fallen world). However, it is a problem for the evolutionists because adaptive changes do not build these impressively complex systems, leaving the evolutionists with no plausible way to account for them.

As I have pointed out elsewhere, natural selection cannot be a major player in accounting for the majority of patterns of adaptation we see in animals around us (Lightner, 2015). It neither accounts for the origin of adaptive diversity, nor is a particularly effective mechanism at fixing beneficial alleles under most circumstances. However, that does not mean that it never plays a role. In the case of the pocket mice, it probably does play at least some role in maintaining the distinct color differences between mice in these neighboring regions.

Several unsubstantiated claims are made in the video. One is that mutations are random. This is a foundational assumption of neo-Darwinism, but it has long been known that mutations tend to occur at hotspots. They are not really random as to when or where they occur (Noble, 2013).

A second unsubstantiated claim is that the mice have no preference for a light or dark background; it is the predators that make all the difference. Mice do occasionally cross the boundaries, and there is gene flow between the light and dark populations (Hoekstra et al., 2005). However, this is insufficient to support this strong assertion. There does not seem to be serious consideration of factors other than natural selection that may have contributed to the pattern seen (e.g., dark mice choosing to migrate into the area with dark rocks resulting in a founder effect; the possibility of non-Mendelian inheritance such as biased gene conversion, etc.).

This brings up a final point. Carroll claims that accounting for this pattern is simple when it comes to the math. He says that if one mouse in 100,000 is born with this black coloration and there are hundreds of thousands born each year, and if the black color gives that mouse a 10% advantage, then it only takes about 100 years for this trait to take over the population. He is ignoring genetic drift and one other big problem. How does this black mouse get born on the right background? One would have needed a sizable population of light mice on the dark background to have the rare mutant born there, yet this seems unlikely. If the dark mouse is born on a light background, it is at a disadvantage. If it does not have a preference for either background, how do we get this mouse to enter the environment for which it is best suited? Something significant appears to be missing from Carroll's explanation.

I very much enjoyed watching the video, especially since coat color genetics is one area of interest for me. In fact, it was my in-depth study of the *mc1r* in other animals that made it increasingly obvious to me that random mutation plus natural selection cannot account for many of the patterns we see (Lightner, 2008). The video is nicely done, but I do not feel it is suitable for showing to students without addressing worldview issues. As is common elsewhere, reasonable scientific conclusions and naturalistic evolutionary philosophy are boldly proclaimed with the same level of certainty. Students need help being able to sort through these issues, but the videos are only helpful for that if there is thought-provoking critical analysis that follows.

References

Hoekstra, H.E., J.G. Krenz, and M.W. Nachman. 2005. Local adaptation in the rock pocket mouse (*Chaetodipus intermedius*): natural selection and phylogenetic history of populations. *Heredity* 94:217–228.

Lightner, J.K. 2008. Genetics of coat color I: the melanocortin 1 receptor (MC1R). *Answers Research Journal* 1:109–116.

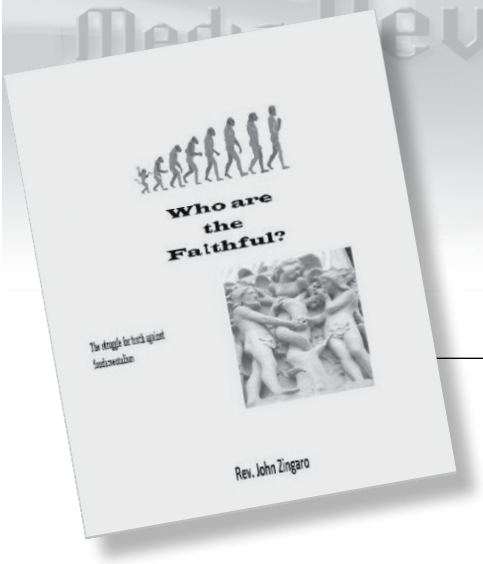
Lightner, J.K. 2015. Natural selection: assessing the role it plays in our world. *Answers Research Journal* 8:111–119.

Nachman, M.W. 2005. The genetic basis of adaptation: lessons from concealing coloration in pocket mice. *Genetica* 123:125–136.

Noble, D. 2013. Physiology is rocking the foundations of evolutionary biology. *Experimental Physiology* 98(8): 1235–1243.

Jean Lightner
jklightner@gmail.com

Media Reviews



Who Are the Faithful? The Struggle for Truth Against Fundamentalism

by John Zingaro

Self-published, San Bernadino, CA, 2012, 406 pages, $30.00

Author Zingaro was reared a Catholic and attended Catholic Schools from kindergarten through eighth grade (p. 8). He claims that not once during his Catholic education did he read the Bible in school, nor was it read in church. He relates that the first time he read a Bible was at age 28 (p. 12). Zingaro began losing his Catholic faith when he was in college studying journalism, and he soon joined a Protestant church. He was also at this time in life a missionary in Africa and active in Bible teaching.

He left Christianity due to the indoctrination in evolution received when a student at the Pittsburgh Presbyterian Theological Seminary. It seemed the most important thing learned at the Presbyterian school was to *not* trust the biblical record. He writes that one day when reading a commentary of "one of my highly respected professors … Old Testament scholar, Dr. Donald Gowan … wrote matter-of-factly in a commentary on Genesis that Adam and Eve had not existed as real individuals. Rather, they were merely symbolic characters representing all of humankind" (p. 29).

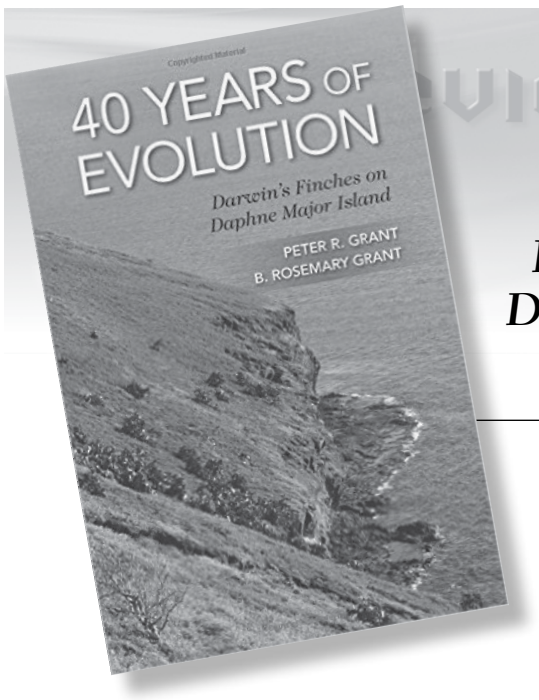
In seminary, Zingaro ended up believing the Bible was not a product of God's revelation but rather was written by people who were "swayed by the prejudices of their day" (p. 128). As a result of this and other instruction in the seminary, his faith was shaken to the point that only a small part of Orthodox Christianity remained. By the time his seminary experience ended, Zingaro writes that he was on his way to becoming a full-fledged Darwinist (p. 128), which is the topic of most of his 406-page book (p. 83). He concludes that the evidence scientists "saw with their own eyes did not match the stories in the Scriptures" (p. 86). His descent into full-fledged Darwinism included accepting the conclusion that "natural Selection is ultimately the creator of all life."

Nonetheless, Zingaro was ordained as a Presbyterian minister in 1994 and has been the pastor of First Presbyterian Church of Newton, New Jersey since 2007. Previous to this he served for 13 years at the Bryn Mawr Presbyterian Church in Cottage Grove, Wisconsin, a suburb of Madison.

Much of this book (pp. 161-379) is about the Dover Intelligent Design trial (which Zingaro incorrectly calls a federal court case) concerning those who take the Bible literally and those who do not (p. 5). He quotes extensively from the trial documents in an effort to show that Darwinism has been proven to be true, and no credible opposition to it exists. He spends much time bemoaning the fact that many people reject what he calls the proven fact of Darwinism (p. 151).

The judge in the Dover case ruled in favor of those who reject the biblical record, thus violating the first amendment of the Constitution, which requires the state to be religiously neutral. Zingaro concludes that the "fundamentalists [creationists] are deluded" (p. 5) and he now totally accepts the current orthodox evolutionist claims—with little or no awareness or understanding of the other side and the problem with many of the evolutionary claims. In short, although a long book, Zingaro gives very few valid scientific arguments to support his conclusions about Darwinism. He gives very little space weighing the pro and con arguments for the views he discusses. Although he had many close female friends, Zingaro later revealed that he accepted the gay lifestyle and contracted colon cancer, a disease that is common among sexually active male homosexuals (p. 135). This book is yet another well-documented example of a lack of training in apologetics, leading to indoctrination into secular humanism and Darwinism.

Jerry Bergman
jerrybergman30@yahoo.com



40 Years of Evolution: Darwin's Finches on Daphne Major Island

by Peter R. Grant and B. Rosemary Grant

Princeton University Press, Princeton, NJ, 2014, 400 pages, $42.00

For 40 years, Peter and Rosemary Grant conducted intensive field research on finches in the Galápagos Islands. While many biologists study populations and make inferences about what happened in the past, the Grants actually individually identified, measured, and followed finches on Daphne Major Island. This type of prospective study gave them a significant advantage in assessing how natural selection and other factors affect a population. The bulk of their work was with the medium ground finch, although they also collected data on several other species. The results of their work changed much of what was previously believed about natural selection and speciation.

The medium ground finch was of primary interest because it was variable in a number of traits, especially beak size. Natural selection can work only if there is variability, so they were a logical choice for research on this topic. The Grants found that natural selection wasn't constant but operated strongly during specific periods of time when the environment changed, namely, during droughts. During the most severe droughts, many birds died, and the size of the beak was the trait that made the difference between life and death. Thus, it can be said that natural selection targeted beak size. As a result, the average beak size changed in the population. The change in the mean of a trait is what the Grants refer to as evolution.

Interestingly, natural selection did not always operate in the same direction. In some years the small seeds were depleted and the death toll was greatest among birds with smaller beaks. However, one drought was preceded by some very wet years that caused the island to be overrun by plants bearing small seeds. In that drought it was birds with larger beaks that were disadvantaged because their food source was depleted first. The reality that natural selection oscillates in direction has some important implications.

The Grants talk about natural selection as a driving force for divergence. And indeed, they explain in detail in the book how natural selection changed the average beak size in the medium ground finches. However, this does not mean that natural selection was helping the birds adapt. In fact, it was reducing useful variety. The food sources returned after the return of the rains, yet many of the birds with a beak size ideal for exploiting the food source had died when natural selection was operating.

Such oscillating weather patterns that underlie natural selection can actually hinder adaptation by putting a population at serious risk. However, the Grants found another factor that affected the average beak size of the birds—hybridization. In many years a small number of individuals from the population bred with members of another species. As their offspring backcrossed with one of the parental species, new alleles were brought into the population, and variation was increased.

Hybridization can have a number of different long-term effects, depending on the specifics involved. In some cases two different species can coalesce, the reverse of speciation. At other times the hybrids may breed among themselves rather than backcross with one of the parent species. In doing so they may become a new species. The Grants found both these patterns among the finches they studied. In the case of the cactus finch, hybridization had a much stronger influence on body size and beak length than selection, which, when detected, was in the opposite direction than the population evolved during their study.

The Grants expand on the significance of hybridization and its relationship to speciation over several chapters. It is not just their work but also a number of other studies that have shown that hybridization is important in the natural history of many species. In fact, it is now believed that hybridization is likely a critical catalyst for adaptive radiations, where organisms rapidly diversify and fill a variety of environmental niches, such as the finches have done in the Galápa-

gos (Abbott et el., 2013). Hybridization appears to allow for the introduction of essential genetic diversity that allows for such adaptive responses to the environment.

The Grants also made some important discoveries about factors associated with a species colonizing a new area. They had noticed large ground finches sometimes visited Daphne Island, but they never stayed to breed until 1982–83, a time when conditions were very favorable. The initial population of two females and three males produced young, but only a brother/sister pair survived to breed. Subsequently, there were signs of inbreeding depression until other immigrants of the species joined them. The large ground finches that contributed to this colonization came from at least three different islands.

One discovery they found particularly surprising was that the birds who stayed to form the breeding colony were genetically different from those that chose to leave before the breeding season. This is in opposition to what biologists often assume. Thus, habitat choice (by the birds) and environmental uncertainties (which may result in natural selection) both played a role in the initial establishment of the population. In chapter 6 the Grants cover various factors, including the founder effect, immigration, and genetic drift, which contribute to the genetic makeup of the new population.

Since the major dry season food source for the large ground finch is also an important food source for the large-beaked medium ground finches, it was expected that competition might end up affecting the birds. Indeed this occurred 22 years after the large ground finch population was established. Their numbers had greatly increased when another drought hit in 2003–2004. There was a very high death toll in both large and medium ground finches. No selection was evident in the large ground finch population, but the larger beaked medium ground finches were more seriously disadvantaged than those with smaller beaks in their population. Since beak size is highly heritable, this was reflected in an unprecedented decrease in average beak size. In later years, when the larger seeds again were plentiful, the medium ground finch population did not recovery the lost variability to enable them to exploit the resource.

The book is an excellent overview of the Grants' 40 years of research. It is well written and well laid out, but there are enough technical details that are relevant that it is not an easy read. There is a helpful summary at the end of each chapter, appendices with extra information, a list of references, and a subject index. Although the authors are evolutionists who hold to the secular old-earth timescale, this is still a valuable book that can provide creationists with important insights into some of the factors affecting the natural history of created kinds.

Reference

Abbott, R., et al. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26(2): 229–246.

Jean Lightner
jklightner@gmail.com

Letters to the Editor

The policy of the editorial staff of* CRSQ *is to allow letters to the editor to express a variety of views. As such, the content of all letters is solely the opinion of the author, and does not necessarily reflect the opinion of the* CRSQ *editorial staff or the Creation Research Society.

This letter ought to have appeared with Andrew Snelling's letter in the Fall 2015 issue. I apologize for this oversight.

—Danny R. Faulkner, Editor

Reply to *CRSQ* Letter to the Editor, "iDINO Corrections"

In CRSQ 52(2):151, Dr. Snelling identified a place name error in our article titled "Radiocarbon in Dinosaur and Other Fossils." In CRSQ 51(4):306 we wrote, "Igneous petrologist Andrew Snelling carbon dated fossil wood extracted from the middle Triassic Hawkesbury Sandstone of Queensland." This should instead have read, "Igneous petrologist Andrew Snelling carbon dated fossil wood extracted from the Newcastle Coal Measures in the Sydney Basin." Incidentally, the result under discussion appears in Figure 6 under "coalified bark." However, this correction does not alter our conclusions.

Brian Thomas and Vance Nelson

Instructions to Authors

Submission

Electronic submissions of all manuscripts and graphics are preferred and should be sent to the editor of the *Creation Research Society Quarterly* in Word, WordPerfect, or Star-Office/Open Office (see the inside front cover for address). Printed copies also are accepted. If submitting a printed copy, an original plus two copies of each manuscript should be sent to the editor. The manuscript and copies will not be returned to authors unless a stamped, self-addressed envelope accompanies submission. If submitting a manuscript electronically, a printed copy is not necessary unless specifically requested by the *Quarterly* editor. Manuscripts containing more than 35 pages (double-spaced and including references, tables, and figure legends) are discouraged. An author who determines that the topic cannot be adequately covered within this number of pages is encouraged to submit separate papers that can be serialized.

All submitted manuscripts will be reviewed by two or more technical referees. However, each section editor of the *Quarterly* has final authority regarding the acceptance of a manuscript for publication. While some manuscripts may be accepted with little or no modification, typically editors will seek specific revisions of the manuscript before acceptance. Authors will then be asked to submit revisions based upon comments made by the referees. In these instances, authors are encouraged to submit a detailed letter explaining changes made in the revision, and, if necessary, give reasons for not incorporating specific changes suggested by the editor or reviewer. If an author believes the rejection of a manuscript was not justified, an appeal may be made to the *Quarterly* editor (details of appeal process at the Society's web site, www.creationresearch.org).

Authors who are unsure of proper English usage should have their manuscripts checked by someone proficient in the English language. Also, authors should endeavor to make certain the manuscript (particularly the references) conforms to the style and format of the *Quarterly*. Manuscripts may be rejected on the basis of poor English or lack of conformity to the proper format.

The *Quarterly* is a journal of original writings, and only under unusual circumstances will previously published material be reprinted. Questions regarding this should be submitted to the Editor (CRSQeditor@creationresearch.org) prior to submitting any previously published material. In addition, manuscripts submitted to the *Quarterly* should not be concurrently submitted to another journal. Violation of this will result in immediate rejection of the submitted manuscript. Also, if an author uses copyrighted photographs or other material, a release from the copyright holder should be submitted.

Appearance

Manuscripts shall be computer-printed or neatly typed. Lines should be double-spaced, including figure legends, table footnotes, and references. All pages should be sequentially numbered. Upon acceptance of the manuscript for publication, an electronic version is requested (Word, WordPerfect, or Star-Office/Open Office), with the graphics in separate electronic files. However, if submission of an electronic final version is not possible for the author, then a cleanly printed or typed copy is acceptable.

Submitted manuscripts should have the following organizational format:

1. Title page. This page should contain the title of the manuscript, the author's name, and all relevant contact information (including mailing address, telephone number, fax number, and e-mail address). If the manuscript is submitted by multiple authors, one author should serve as the corresponding author, and this should be noted on the title page.

2. Abstract page. This is page 1 of the manuscript, and should contain the article title at the top, followed by the abstract for the article. Abstracts should be between 100 and 250 words in length and present an overview of the material discussed in the article, including all major conclusions. Use of abbreviations and references in the abstract should be avoided. This page should also contain at least five key words appropriate for identifying this article via a computer search.

3. Introduction. The introduction should provide sufficient background information to allow the reader to understand the relevance and significance of the article for creation science.

4. Body of the text. Two types of headings are typically used by the *CRSQ*. A major heading consists of a large font bold print that is centered in column, and is used for each major change of focus or topic. A minor heading consists of a regular font bold print that is flush to the left margin, and is used following a major heading and helps to organize points within each major topic. Do not split words with hyphens, or use all capital letters for any words. Also, do not use bold type, except for headings (italics can be occasionally used to draw distinction to specific words). Italics should not be used for foreign words in common usage, e.g., "et al.", "ibid.", "ca." and "ad infinitum." Previously published literature should be cited using the author's last name(s) and the year of publication (ex. Smith, 2003; Smith and Jones, 2003). If the citation has more than two authors, only the first author's name should appear (ex. Smith et al., 2003). Contributing authors should examine this issue of the *CRSQ* or consult the Society's web site for specific examples as well as a more detailed explanation of manuscript preparation. Frequently-used terms can be abbrevi-

ated by placing abbreviations in parentheses following the first usage of the term in the text, for example, polyacrylamide gel electrophoresis (PAGE) or catastrophic plate tectonics (CPT). Only the abbreviation need be used afterward. If numerous abbreviations are used, authors should consider providing a list of abbreviations. Also, because of the variable usage of the terms "microevolution" and "macroevolution," authors should clearly define how they are specifically using these terms. Use of the term "creationism" should be avoided. All figures and tables should be cited in the body of the text, and be numbered in the sequential order that they appear in the text (figures and tables are numbered separately with Arabic and Roman numerals, respectively).

5. Summary. A summary paragraph(s) is often useful for readers. The summary should provide the reader an overview of the material just presented, and often helps the reader to summarize the salient points and conclusions the author has made throughout the text.

6. References. Authors should take extra measures to be certain that all references cited within the text are documented in the reference section. These references should be formatted in the current *CRSQ* style. (When the *Quarterly* appears in the references multiple times, then an abbreviation to *CRSQ* is acceptable.) The examples below cover the most common types of references:

Robinson, D.A., and D.P. Cavanaugh. 1998. A quantitative approach to baraminology with examples from the catarrhine primates. *CRSQ* 34:196–208.

Lipman, E.A., B. Schuler, O. Bakajin, and W.A. Eaton. 2003. Single-molecule measurement of protein folding kinetics. *Science* 301:1233–1235.

Margulis, L. 1971a. The origin of plant and animal cells. *American Scientific* 59:230–235.

Margulis, L. 1971b. *Origin of Eukaryotic Cells*. Yale University Press, New Haven, CT.

Hitchcock, A.S. 1971. *Manual of Grasses of the United States*. Dover Publications, New York, NY.

Walker, T.B. 1994. A biblical geologic model. In Walsh, R.E. (editor), *Proceedings of the Third International Conference on Creationism* (technical symposium sessions), pp. 581–592. Creation Science Fellowship, Pittsburgh, PA.

7. Tables. All tables cited in the text should be individually placed in numerical order following the reference section, and not embedded in the text. Each table should have a header statement that serves as a title for that table (see a current issue of the *Quarterly* for specific examples). Use tabs, rather than multiple spaces, in aligning columns within a table. Tables should be composed with *14-point type* to insure proper appearance in the columns of the *CRSQ*.

8. Figures. All figures cited in the text should be individually placed in numerical order, and placed after the tables. Do not embed figures in the text. Each figure should contain a legend that provides sufficient description to enable the reader to understand the basic concepts of the figure without needing to refer to the text. Legends should be on a separate page from the figure. All figures and drawings should be of high quality (hand-drawn illustrations and lettering should be professionally done). Images are to be a minimum resolution of 300 dpi at 100% size. Patterns, not shading, should be used to distinguish areas within graphs or other figures. Unacceptable illustrations will result in rejection of the manuscript. Authors are also strongly encouraged to submit an electronic version (.cdr, .cpt, .gif, .jpg, and .tif formats) of all figures in individual files that are separate from the electronic file containing the text and tables.

Special Sections

Letters to the Editor:

Submission of letters regarding topics relevant to the Society or creation science is encouraged. Submission of letters commenting upon articles published in the *Quarterly* will be published two issues after the article's original publication date. Authors will be given an opportunity for a concurrent response. No further letters referring to a specific *Quarterly* article will be published. Following this period, individuals who desire to write additional responses/comments (particularly critical comments) regarding a specific *Quarterly* article are encouraged to submit their own articles to the *Quarterly* for review and publication.

Editor's Forum:

Occasionally, the editor will invite individuals to submit differing opinions on specific topics relevant to the *Quarterly*. Each author will have opportunity to present a position paper (2000 words), and one response (1000 words) to the differing position paper. In all matters, the editor will have final and complete editorial control. Topics for these forums will be solely at the editor's discretion, but suggestions of topics are welcome.

Book Reviews:

All book reviews should be submitted to the book review editor, who will determine the acceptability of each submitted review. Book reviews should be limited to 1000 words. Following the style of reviews printed in this issue, all book reviews should contain the following information: book title, author, publisher, publication date, number of pages, and retail cost. Reviews should endeavor to present the salient points of the book that are relevant to the issues of creation/evolution. Typically, such points are accompanied by the reviewer's analysis of the book's content, clarity, and relevance to the creation issue.

Creation Research Society Membership/Subscription Application and Renewal Form

The membership/subscription categories are defined below:

1. **Voting Member** .....Those having at least an earned master's degree in a recognized area of science.
2. **Sustaining Member** .....Those without an advanced degree in science, but who are interested in and support the work of the Society.
3. **Student Member** .....Those who are enrolled full time in high schools, undergraduate colleges, or postgraduate science programs (e.g., MS, PhD, MD, and DVM). Those holding post-doctoral positions are not eligible. A graduate student with a MS degree may request voting member status while enrolled as a student member.
4. **Senior Member** .....Voting or sustaining members who are age 65 or older.
5. **Life Member** .....A special category for voting and sustaining members, entitling them to a lifetime membership in the Society.
6. **Subscriber** .....Libraries, churches, schools, etc., and individuals who do not subscribe to the Statement of Belief.

All members (categories 1–5 above) must subscribe to the Statement of Belief as defined on the next page.

Please complete the lower portion of this form and mail it with payment to CRS Membership Secretary, 6801 N. Highway 89, Chino Valley, AZ 86323, or fax for credit card payment to (928) 636-1153. Applications may also be completed online at creationresearch.org.

✂ ----------

This is a ❑ new ❑ renewal application for the subscription year beginning Summer ❑ 2014 ❑ _____. *(Please type or print legibly.)*

Name _____ Address _____

City _____ State _____ Postal/Zip code _____ Country _____

Phone (optional) _____ Email _____

Degree _____ Field _____

Year granted _____ Institution _____

Presently associated with _____

I have read and subscribe to the CRS Statement of Belief. Signature _____

For foreign orders, including Canadian, payment must be made in U.S. dollars by a check drawn on a U.S. bank, international money order, or credit card. *Please do not send cash.*

Indicate applicable category ⬇	*Indicate payment* ⬇			
	Paper**			
❑ Voting ❑ Sustaining	USA	Canada Mexico	Other countries	**Paper-less**‡
❑ Regular [per year]	❑ $38	❑ $58	❑ $75	❑ $31
❑ Senior [per year]	❑ $33	❑ $53	❑ $70	❑ $26
❑ Life member	❑ $500	❑ $500	❑ $500	❑ $500
❑ Student* [per year]	❑ $33	❑ $53	❑ $70	❑ $26
❑ Subscriber [per year]	❑ $41	❑ $61	❑ $78	❑ $34

* Student members are required to complete the bottom portion of this form.
NOTE: Student members may qualify for the *Future Leaders Sponsorship* program. See the CRS website at www.creationresearch.org for details.
** Rates for the paper option include postage for First Class Mail International

‡ **PAPERLESS option:** You may opt out of receiving paper copies of the CRS periodicals (*CRS Quarterly* and *Creation Matters*). By choosing this option you may register for access to the Premium Area of the website, where you may view or download electronic (PDF) versions of these publications. Of course, regular members and subscribers may also have access to the Premium Area. Only members, however, will have access to the Members Exclusive Area of the website.

Member/Subscriber	$_____ per year
	x _____ years
SUBTOTAL	$_____
Optional contribution	+ $_____
Life membership	+ $_____
TOTAL	$_____

❑ Visa ❑ MasterCard ❑ Discover
❑ American Express ❑ Check/money order

Card number _____
Expiration date (mo/yr) _____
Phone number (_____) _____
Signature _____

***Student Members* are required to complete the following:**

School or institution now attending _____

_____

Your current student status: ❑ high school; ❑ undergraduate; graduate program ❑ MS ❑ PhD; ❑ other _____

Year you expect to graduate or complete your degree _____

Major, if college or graduate student _____

Signature _____

Order Blank for Past Issues

Cost of complete volumes (per volume): .....members (all categories) – $18.00 + S/H
nonmembers and subscribers (libraries, schools, churches, etc.) – $25.00 + S/H

Cost of single issues (per issue):.....members (all categories) – $5.00 + S/H
nonmembers and subscribers (libraries, schools, churches, etc.) – $7.00 + S/H

	Number					Number				
Volume	1	2	3	4	Volume	1	2	3	4
21	☐	☐	☐	☐	37	☐	☐	☐	☐
22	☐	☐	☐	☐	38	☐	☐	☐	☐
23	☐	☐	☐	☐	39	☐	☐	☐	☐
24	☐	☐	☐	☐	40	☐	☐	☐	☐
25	☐	☐	☐	☐	41	☐	☐	☐	☐
26	☐	☐	☐	☐	42	☐	☐	☐	☐
27	☐	☐	☐	☐	43	☐	☐	☐	☐
28	☐	☐	☐	☐	44	☐	☐	☐	☐
29	☐	☐	☐	☐	45	☐	☐	☐	☐
30	☐	☐	☐	☐	46	☐	☐	☐	☐
31	☐	☐	☐	☐	47	☐	☐	☐	☐
32	☐	☐	☐	☐	48	☐	☐	☐	☐
33	☐	☐	☐	☐	49	☐	☐	☐	☐
34	☐	☐	☐	☐	50	☐	☐	☐	☐
35	☐	☐	☐	☐	51	☐	☐	☐	☐
36	☐	☐	☐	☐	52	☐	☐	☐	☐

Add 20% for postage (**for U.S. orders:** min. $6, max. $18; **for Canadian orders:** min. $10, no max.; **for other foreign orders:** min. $15, no max.) **Total enclosed:** $_____

Make check or money order payable to Creation Research Society. Please do not send cash. For foreign orders, including Canadian, please use a check in U.S. funds drawn on a U.S. bank, an international money order, or a credit card.

(Please type or print legibly)

Name _____ Address _____

City _____ State _____ Zip _____ Country _____

❏ Visa ❏ MasterCard ❏ Discover ❏ American Express Card number _____

Expiration date (mo/yr) _____ Signature _____

Mail to: Creation Research Society, 6801 N. Highway 89, Chino Valley, AZ 86323, USA

Creation Research Society

History—The Creation Research Society was organized in 1963, with Dr. Walter E. Lammerts as first president and editor of a quarterly publication. Initially started as an informal committee of 10 scientists, it has grown rapidly, evidently filling a need for an association devoted to research and publication in the field of scientific creation, with a current membership of over 600 voting members (graduate degrees in science) and about 1000 non-voting members. The *Creation Research Society Quarterly* is a peer-reviewed technical journal. It has been gradually enlarged and modified, and is currently recognized as one of the outstanding publications in the field. In 1996 the *CRSQ* was joined by the newsletter *Creation Matters* as a source of information of interest to creationists.

Activities—The Society is a research and publication society, and also engages in various meetings and promotional activities. There is no affiliation with any other scientific or religious organizations. Its members conduct research on problems related to its purposes, and a research fund and research center are maintained to assist in such projects. Contributions to the research fund for these purposes are tax deductible. As part of its vigorous research and field study programs, the Society operates The Van Andel Creation Research Center in Chino Valley, Arizona.

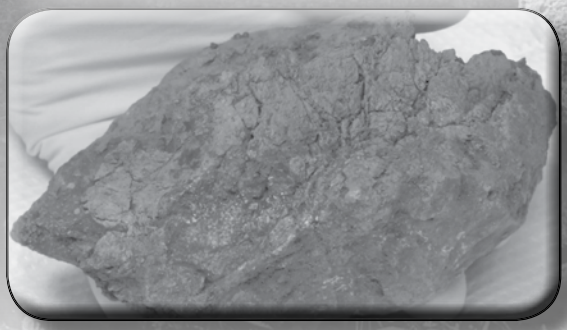
Membership—Voting membership is limited to scientists who have at least an earned graduate degree in a natural or applied science and subscribe to the Statement of Belief. Sustaining membership is available for those who do not meet the academic criterion for voting membership, but do subscribe to the Statement of Belief.

Statement of Belief—Members of the Creation Research Society, which include research scientists representing various fields of scientific inquiry, are committed to full belief in the biblical record of creation and early history, and thus to a concept of dynamic special creation (as opposed to evolution) both of the universe and the earth with its complexity of living forms. We propose to re-evaluate science from this viewpoint, and since 1964 have published a quarterly of research articles in this field. *All members of the Society subscribe to the following statement of belief:*

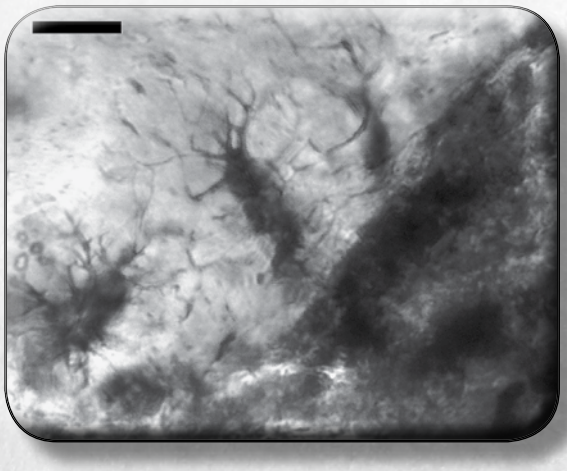
1. The Bible is the written Word of God, and because it is inspired throughout, all its assertions are historically and scientifically true in all the original autographs. To the student of nature this means that the account of origins in Genesis is a factual presentation of simple historical truths.
2. All basic types of living things, including humans, were made by direct creative acts of God during the Creation Week described in Genesis. Whatever biological changes have occurred since Creation Week have accomplished only changes within the original created kinds.
3. The Great Flood described in Genesis, commonly referred to as the Noachian Flood, was a historical event worldwide in its extent and effect.
4. We are an organization of Christian men and women of science who accept Jesus Christ as our Lord and Savior. The act of the special creation of Adam and Eve as one man and woman and their subsequent fall into sin is the basis for our belief in the necessity of a Savior for all people. Therefore, salvation can come only through accepting Jesus Christ as our Savior.

iDINO II

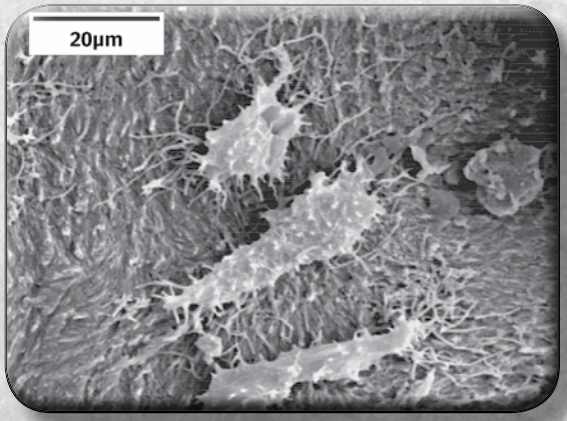
Investigation of Dinosaur Intact Natural Osteo-tissue



A fragment of the *Triceratops* brow horn. Fragments, such as this one, still contain tissue and cells.



Microscopic examination of tissue extracted from a *Triceratops* horn reveals bone cells still present.



Electron microscope picture of intact bone cells still in tissue extracted from a *Triceratops* horn.

How can pliable, stretchable tissue survive inside dinosaur fossils for over 65 million years?

How can this tissue still contain intact cells and even dinosaur proteins?

How can this fragile biological material survive for so long?

The answer to these questions directly challenges the current, evolutionary-biased, geologic timescale.

The Creation Research Society began its iDINO research initiative for the purpose of studying soft tissue in dinosaur fossils. The first phase of the project detected pliable, unfossilized tissue in a brow horn of a *Triceratops*. Within this tissue were intact osteocytes (bone cells). Some results from the iDINO project have been published in a technical microscopy journal and presented at an international microscopy conference. The Spring 2015 issue of the *Creation Research Society Quarterly* also features a special report of the iDINO project. Plus, to further spread the important information about soft tissue, the Society is developing a video (Echoes of the Jurassic).

The **second phase** of the project (iDINO II) will look more extensively at the process of tissue preservation. Evolutionists have offered various theories of how this tissue could survive for millions of years. iDINO II will methodically investigate these preservation claims, assessing their plausibility.

The iDINO results have already provided a strong challenge to the evolutionary worldview. More extensive and detailed examination may provide even stronger evidence that the age of dinosaur fossils is far less than 65 million years. To this end, the Society continues to seek those willing to fund this project with either one-time gifts or monthly donations.

For more information contact us at (928) 636-1153 or crsvarc@crsvarc.com.

Also visit http://tinyurl.com/nphm2c4 for project updates and details.