# Random Worms: Evidence of Random and Nonrandom Processes in the Chromosomal Structure of Archaea, Bacteria and Eukaryotes

Glenn R. Morton and Gordon Simons

One of the central debates in Christian apologetics concerns the role of chance and randomness within living systems and the presumed incompatibility of chance with complex organisms containing high informational content. In order to address this issue, the chromosomal organization of genes for ten different species of varying levels of complexity was examined for evidence of randomness in the gene structure. The results show an interplay of random and nonrandom processes with the more complex eukaryotes evidencing more randomness in their genetic structure than is found in the simpler prokaryotic organisms: bacteria and archaea. While almost all anti-evolutionary views reject any role for chance or randomness in biology, we find that the Bible supports a much more compatible perspective.

Glenn Morton

Gordon Simons
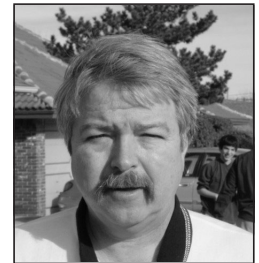
Biology and chance

One of the philosophical issues about which Christians have debated over the past century concerns the role of chance in the biological realm. Many Christians have rejected any role for chance. The conservative creationist Henry Morris states: "Chance and design are antithetical concepts."[1]

It is not only Christians opposed to evolution who reject chance in the biological realm. Rejection of chance in the biological realm is a common trait across religious and theological boundaries. The Jewish anti-evolutionist, Lee Spetner writes: "The information required for large-scale evolution cannot come from random variations."[2] (While noting that this second quote uses the word *random* rather than *chance*, we will not attempt to infer whether the author meant to convey a distinction. A discussion of terminology appears later.)

A member of the Reunification Church, Jonathan Wells wrote: "Furthermore, the integrated complexity of developmental programs cannot plausibly be attributed to chance."[3]

And even Islamic anti-evolutionists take the same position:

Laboratory experiments and probabilistic calculations have definitely made it clear that the amino acids from which life arises cannot have been formed by chance. The cell, which supposedly emerged by chance under primitive and uncontrolled terrestrial conditions according to evolutionists, still cannot be synthesized even in the most sophisticated, high-tech laboratories of the 20th century.[4]
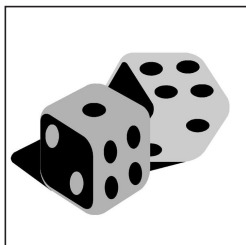
A major objective of the Intelligent Design movement has been to show that chance cannot work. Dembski states: "Now a little reflection makes clear that a pattern need not be given prior to an event to eliminate chance and implicate design."[5]

Glenn Morton, *an ASA member, has a B.S. in Physics from Oklahoma University and works as a geophysicist in the oil industry. He has been an independent consultant and served as manager of geophysical training, as chief geophysicist for China, and as the geophysical manager for the U.S. Offshore. He currently is serving as geophysical manager of the North Sea. Besides publishing articles on geophysics, Glenn has published over fifty articles in the area of creation and evolution along with four books on the topic. He and Debi have three grown sons. Correspondence may be sent to him at: gmorton@kmg.com.*

Gordon Simons *is a faculty member at the University of North Carolina in Chapel Hill, where he is a professor of statistics, and has twice served as the chairman of the Department of Statistics. He earned his Ph.D. in statistics at the University of Minnesota in 1966. Gordon has published numerous papers concerned with probability and statistics, mostly of a theoretical nature. While he has been interested in the relationship between science and Christianity, and a member of ASA for many years, this is his first professional effort of this kind. He and his wife, Karen, have three grown daughters.*

God ordered or allowed the use of [chance] at critical places in the biblical history. If God is incompatible with chance in his dealings with this world, it seems odd that he allowed and commanded the use of such systems.

There are many other Christians of varying theological persuasions who reject the role of chance in biology.[6] Indeed, if this position is not the majority position in conservative Protestant theology, then it is very close to it.

Finally, we observe that Christians appear much more troubled by assertions of chance in biology than by chance in the nonlife sciences, for instance, in physics with the decay of nuclei.

The Bible and Chance

One of the difficulties raised by the rejection of chance in nature lies in the fact that God ordered or allowed the use of such systems at critical places in the biblical history. If God is incompatible with chance in his dealings with this world, it seems odd that he allowed and commanded the use of such systems. The *Urim and Thrummim* which the priest carried is widely believed to have been a tool for casting lots before the Lord.[7] The Hebrews believed what Prov. 16:33 says: "The lot is cast into the lap, but every decision is from the Lord." Proverbs 18:18 would indicate that the Jews thought God was the true decision maker when chance was involved. That verse says: "Casting the lot settles disputes and keeps strong opponents apart." In 1 Chron. 24:1–5, 1 Chron. 24:31, and 1 Chron 25:8, David cast lots to determine the order of the service for the sanctuary officials. God used the chance lots of the sailors to identify Jonah as the source of their troubles (Jon. 1:7). In Lev. 16:8, God told the Israelites to cast lots for the sacrificial goat. God told Joshua to cast lots in order to identify Achan, the guilty keeper of the Canaanite booty. In Josh. 18:8, we see Joshua casting lots for the assignment of land to the various tribes. In Acts 1:24–26, the disciples used chance, the casting of lots, to determine who should take over the apostolic ministry of Judas. Because of the biblically widespread use of chance to determine God's will, it is truly amazing that many modern Christians reject chance in biology as being totally incompatible with God's control.

If God cannot control chance, how can he control the lots above? God predetermined the result yet used a tool of chance. If God cannot use chance, then one must logically conclude that God did not foreknow how the land would be divided among the tribes,

that God did not foreknow that Jonah would be picked, that God did not foreknow that Achan was the one who would be chosen or that Matthias would step into the apostolic line. This is a position which basically says that God is not omnipotent or omniscient. If God can use chance in his dealings with Israel and the early church, then why do we say he has no ability to use chance in biology? God can, has, and does control the stochastic process even if we do not understand how it happens.

One thing Christians must keep in mind is that our perspective on chance is not God's. Humans are not always able to distinguish between appearance of chance and the actuality of chance. But we cannot say that God is equally so limited. The molecules in a gas move according to deterministic laws, but the Maxwellian distribution of their velocities gives the appearance of chance. On the other hand, quantum phenomena appear to be the actuality of chance. But that might not be the view from God's perspective given the biblical references above. In biology, we see the fertilization of an egg as the result of a random or nearly random event. A single sperm may have only a one in fifty billion chance of being the "lucky" winner of the race to the egg. Yet God proclaims through Jeremiah: "Before I formed you in the womb I knew you, before you were born I set you apart" (1:5, NIV). If God were unable to control chance, why would he make such a statement?

Biological Evidence for Chance?

In the past, it has been difficult to actually test the chance hypothesis in biological systems. The discussion has revolved too often around debatable probability arguments. The standard argument says that there are too many possible combinations in proteins, or too many possible combinations in DNA/RNA for working sequences to be found by random mutation. These arguments are based on the assumption that very few sequences out of the entire ensemble of possibilities would be capable of performing the sought for task. Examples are legion but Gange provides a good one. He says:

> Hemoglobin contains two trains totaling 574 cars—each selected from among twenty kinds of amino acids. The num-

ber of ways we can assemble these hemoglobin trains is so vast that it is a trillion trillion (repeat twenty times more) times the entire number of stars in the universe, despite this, only one combination known to man carries oxygen most efficiently in your blood.[8]

The weakness of this argument is that Gange cannot prove the last phrase. How can he *know* that only one combination carries oxygen most efficiently? How can he *know* that hemoglobin is it? Since he has no way of comparing the efficiency of hemoglobin against all other possible molecules of that length, much less comparing it to molecules of shorter and longer length, his argument rests upon an untested assumption. On the evolutionary side, this argument is difficult to counter because, like Gange, one too cannot search and find a more efficient molecule so the argument boils down to an opinion about what is unobserved and what is unknown about biological molecules. One can have an opinion about such matters, but neither side can say much worth listening to scientifically.

Genome-sequencing projects provide a way to look at the genetic organization of various organisms … These data provide an excellent platform from which to examine the role of randomness.

Is there a way to break out of this opinion-dominated trap? We believe there is, and the genome-sequencing projects provide a way to look at the genetic organization of various organisms. The genomes of hundreds of organisms have been sequenced in various mapping projects. These data provide an excellent platform from which to examine the role of randomness. If random processes have been active in the genome, then the structure of the chromosome should be consistent with what is expected from a random process. If there is no evidence for random processes and the genome is organized via nonrandom processes, then predictions from nonrandom models of the genome should be possible. In this article, we will test the role of randomness and nonrandomness by looking at the organization of the genes along the chromosomes.

We have examined the genetic organization for six bacteria from five species, two species of the archaea, and a plasmid found in *E. coli*. These organisms in general have one circular chromosome, the only exception being *Vibrio cholerae* which has two chromosomes, the second of which originally may have been a plasmid. At a more complex level, we also have examined the sixteen chromosomes of yeast; the eight chromosomes of the fruitfly, *Drosophila*

melanogaster; and the six chromosomes of the nematode, *Caenorabditusus elegans*. We also examined but rejected as too incomplete at this time, the human genome. The results of our study show some interesting features of genetic organization relating to the central philosophical question of this article—What is the role of chance in biology?

Definitions

Here we will discuss *chance* and *randomness* in their varying usages.

It seems plausible to us that the widespread rejection of chance (and randomness) within the biological realm among various religious peoples is due to a widespread perception that the word chance rules out God as the causative agent—that it leaves no room for any understanding of intelligent design. In contrast, we have observed that the Bible conveys no such concern: chance mechanisms are fully under the sovereign control of God. Or, at the very least, they are never at odds with his permissive intents.

Russian mathematician A. N. Kolmogorov, in 1933, is credited with providing the first axiomatic definition of probability—a definition precise enough to gain the widespread acceptance of mathematicians, yet comprehensive enough to be applicable to a wide range of phenomena.[9] His definition of probability places on a firm foundation the notions of *chance*, *randomness*, *random variables, random processes*, and a variety of related concepts, making possible a rigorous development of the subject of probability.

Mathematical treatments of probability date back to the seventeenth century when French mathematicians Blaise Pascal and Pierre de Fermat analyzed various questions of gaming and gambling. Over the intervening centuries, the subject has engaged the serious attention of many well-known scientists and mathematicians, among them Huygens, Jacob Bernoulli, Abraham de Moive, Pierre de Laplace, Chebyshev, Markov and von Mises. Also, the subject of statistics uses probability theory at its foundation. Early uses of statistics to model biological phenomena trace back to Gregor Mendel[10] and, more recently, to R. A. Fisher, who published widely on such subjects as eugenics, Mendelian inheritance, "Darwinian Evolution by Mutation," "The Evolution of Dominance," and much more. Moreover, he contributed widely, in fundamental ways, to the development of statistical tools and concepts.

So presently there is in place a widely applicable set of mathematical and statistical tools and concepts for modeling and analyzing biological structures and phenomena from a probabilistic perspective. However, a serious lack of understanding of these tools and concepts among nonscientists exists. In particular, most Christians lack a clear conceptual understanding of chance. Still, we believe it safe to assert that most Christians do have a fairly
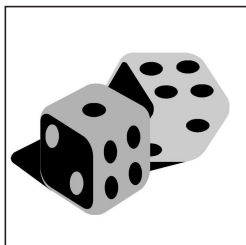
Article
*Random Worms: Evidence of Random and Nonrandom Processes in the
Chromosomal Structure of Archaea, Bacteria and Eukaryotes*

For the purposes of this article, we will opt for an intuitive definition of randomness. In particular, we will take random to mean a process relating to or being defined by events with a particular probability distribution. In this definition, there is no theological implication at all.

well-developed perception of what can be summarized by the phrase "chance mechanism" (such as dice, the roulette wheel, and coin tossing). Below, we will appeal to this intuitive understanding.

For the purposes of this article, we will opt for an intuitive definition of randomness. In particular, we will take *random* to mean a process relating to or being defined by events with a particular probability distribution. In this definition, there is no theological implication at all. We think this is an important point to realize about probability in the theological context. Saying that something has a definite probability of occurrence is not saying that something is totally unpredictable. Probabilities allow the prediction of the system behavior for a large number of iterations. If I have a coin, I can predict with a great deal of certainty that after two million flips of the coin, there will be very close to a 50-50 split between heads and tails. Probability and prediction are not totally incompatible.

We must place another caveat on what is meant by randomness. Randomness can only be measured against a model of randomness, i.e., a stochastic process or system. Such a process or system is inherently a mathematical "recipe" for manipulating the random occurrences and producing an output. The output may be determined from the input random events only after a complex calculation. To start simply, a coin is a simple 2-state stochastic (probabilistic) system. The die used on board games is a 6-state stochastic system. In these two examples, the probability of occurrence for each of the different states is identical, ½ and 1/6 respectively. But that does not have to be the case. One can manipulate the probabilities on a die and have an unequal probability among the six choices. These are termed "loaded dice." Loaded dice will output a sequence of numbers that appears nonrandom if compared to the normal unloaded dice.

Other stochastic systems can be even more complex. The chance of a particular outcome might depend upon the state of affairs at the time the die is rolled. These are called Markov chains. The probability of an outcome depends upon the system state that exists *at the time*. While not subject to random chance, we do see an illustration of

this type of behavior in languages. If, in a sequence of English letters, the letter q is the current state of affairs, one can be nearly 100% certain that the next letter is a u. If one counts the symbols, one will generally find that e is a more common (higher probability) letter than z. Cryptologists use such frequency analysis to decipher coded texts. They arrive at a probabilistic model of what letters are intended by the coder. Probability models are also used in computer networks to avoid bottlenecks. Therefore, probabilistic models are not incompatible with engineering design.

How does one decide if a given sequence is random or influenced by a stochastic system? First, randomness is something that cannot be proven. After comparing a sequence of numbers with a given stochastic process, all one can ever really say is that the sequence is consistent with it being randomly generated by a particular stochastic process or that it is inconsistent with such a process. Secondly, if one is trying to decide if the numerical sequence 2,2,1,2,1,1,1,2 … is random one needs to be sure that this sequence is the output of a 2-state stochastic system. If such a sequence was the outcome of a 6-sided die, the outcome is entirely *nonrandom when compared to an equal probability six-state system*. But if one has a Markov chain with heavy probability weighting for 1's and 2's, if the current state is a 1 or 2, then this output is entirely consistent with such a stochastic process. Thus, when examining the patterns seen in the gene strings, we need to examine several different probability models.

Chromosomal Organization

A chromosome consists of four nucleotides laid out in a double helix. This is the lowest level of organization for the chromosome. There are higher levels. Sequences of nucleotides are functionally connected forming a gene. Genes are systems of nucleotides which perform the function of providing information for the construction of proteins. This is not the end of the organization seen in the genome. At a still higher level of organization, the genes are strung out along the chromosome in groups of genes which all have the same transcription direction. The transcription direction is the direction in which the cellular machinery must read the

gene in order to recover the proper protein information. There are two complementary strands of DNA in organisms, and genes can be found on either strand. Genes found on one strand are transcribed in one direction along the chromosome, and genes found on the opposite strand are transcribed in the other direction. Groups of genes with the same transcription direction are called *gene clusters* or *strings*. We will use the latter term, and to our knowledge this term is new.

A string is merely the consecutive genes transcribed in the same direction. It is the strings which will form the basis for our analysis. Intuitively, to a nonbiologist, it would seem that it really should not matter which direction a gene is transcribed. As we will see, this is not what is observed in the strings. Direction does matter and the amount it matters depends upon the organism.

As noted above, a gene lies along a chromosome and the biochemical machinery reads the nucleotide sequence of the gene and eventually translates that information into a protein. The gene can be read only in one direction by the machinery. But that direction is not constant for all genes on a given chromosome. Approximately half of the genes are transcribed in one direction and the other half in the opposite direction. For instance, *M. genitalium* has 297 genes which must be transcribed in the positive direction and 225 which must be transcribed oppositely in the negative direction. These genes are organized into 86 different strings—contiguous genes which must be transcribed in the same direction at the same time. String length is merely the number of genes in a string. Figure 1 below shows the string lengths for *M. genitalium*, the organism that, until recently, was the shortest known genome.

When one lists the sequence of numbers which represent the number of genes in a string, a question presents itself of immense philosophical importance to the issue of chance and randomness in biology. Is this sequence of numbers random or not? If it is random, then it would be clear evidence of random or chance processes occurring in biology and would directly speak to the issues raised in this article. If they are not random, but are ordered, then they would support the claims of the apologists that chance and random processes are not involved in biological organisms.

Why do gene strings exist? In the process of converting the information contained in DNA into proteins, a gene is first copied into mRNA which then in turn is read by the ribosome which produces the protein. If many mRNA copies of a gene are made, the ribosome will produce many copies of the protein. Two genes that are next to each other on the same DNA strand are transcribed into mRNA together. In these cases, the ribosome receives an mRNA molecule encoding for several proteins. The ribosome will recognize each mRNA separately and produce the correct protein. This is a useful procedure as it allows genes which are linked to the same biochemical processes to be transcribed together and allows the cell to operate efficiently. If several enzymes are required for a given reaction, this procedure ensures that all the proteins are created together. As we will see, this procedure only applies to the bacteria and archaea and not to the "higher" eukaryotes.

The Probability Models Compared with Gene Strings

We downloaded the huge flat files for the organisms listed in Table 1 from the National Center for Biotechnical Information and extracted the relevant gene string data arranged in chromosomes.[11] In the case of *D. melanogaster*, the full genome is not sequenced. The gene containing regions have been sequenced but regions of repeats have not. Thus, they are listed in Table 1 in scaffolds, which are regions of the contiguously sequenced data. It appears that groups sequencing *D. melanogaster* will maintain this format for the final version.

Using standard statistical techniques,[12] three different stochastic models were compared with the gene data. The reader is referred to Hutchison, et al. for the technical details of each model.[13] We will describe each of the models in nontechnical terms hoping to give the reader a basic understanding of the issues. As mentioned above, randomness can only be ascertained by comparing the sequence of numbers with a stochastic model. If a given model does not produce a sequence that matches the observed string size distribution, then the model is wrong and another model must be found.

Model 1 is the simplest. In this model, we compared the chromosomal organization with a statistical model which assumes a probability distribution similar to that of a coin-

Strings 1–20	10	5	15	3	2	1	3	2	8	1	5	8	1	2	14	1	29	1	19	1
Strings 21–40	5	2	1	1	66	1	3	1	2	1	2	3	20	1	20	1	6	9	9	1
Strings 41–60	5	13	1	9	6	1	3	19	1	11	2	9	1	14	3	8	3	2	3	1
Strings 61–80	2	1	1	3	2	1	3	1	3	2	3	1	1	26	1	14	4	4	5	13
Strings 81–86	2	1	1	4	1	11														

Figure 1. String length in *M. Genitalium*, the numbers represent the numbers of genes in a row with the same orientation.
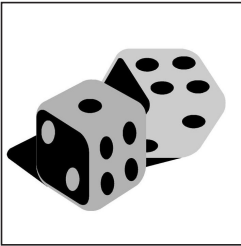
Random Worms: Evidence of Random and Nonrandom Processes in the Chromosomal Structure of Archaea, Bacteria and Eukaryotes

In Table 1 we list the number of genes per string (n/s) and then the 3 standard deviation range for each entry. When we compare the string lengths observed with the expected range for the coin-flip model, we see significant deviations from that in Table 1 among the archaea and bacteria. … The clear conclusion is that this random model totally fails to account for the bacterial data.

Genome	Model 1 Strain	Domain	# Genes n	# Strings s	 n/s	3 Standard Deviation Range
Mycoplasma genitalium	G-37	Bacteria	522	86	6.07	1.73-2.26
Bacillus subtilis	168	Bacteria	4222	1092	3.78	1.91–2.09
Escherichia coli	K12	Bacteria	4405	1292	3.41	1.91–2.09
Escherichia coli	157:H7	Bacteria	5410	1472	3.68	1.91–2.08
Vibrio cholerae 1	N16961	Bacteria	2769	876	3.16	1.88–2.11
Vibrio cholerae 2	N16961	Bacteria	1115	398	2.80	1.82–2.17
Aeropyrum pernix	K1	Archaea	2746	1612	1.70	1.88–2.11
Archaeoglobus fulgidus	DSM4304	Archaea	2486	738	3.37	1.87–2.12
Campylobacter jejuni		Bacteria	4102	1094	3.75	1.91–2.09
Plasmid F		Bacteria	111	32	3.47	1.43–2.57
	Chromosome					
Saccharomyces cerevisiae	1	Eukaryote	112	62	1.80	1.43–2.57
Saccharomyces cerevisiae	2	Eukaryote	446	237	1.88	1.72–2.28
Saccharomyces cerevisiae	3	Eukaryote	186	84	2.21	1.56–2.44
Saccharomyces cerevisiae	4	Eukaryote	856	467	1.83	1.79–2.21
Saccharomyces cerevisiae	5	Eukaryote	309	151	2.04	1.66–2.34
Saccharomyces cerevisiae	6	Eukaryote	145	72	2.01	1.50–2.50
Saccharomyces cerevisiae	7	Eukaryote	609	319	1.91	1.76–2.24
Saccharomyces cerevisiae	8	Eukaryote	295	149	1.97	1.65–2.35
Saccharomyces cerevisiae	9	Eukaryote	233	126	1.84	1.61–2.39
Saccharomyces cerevisiae	10	Eukaryote	416	220	1.89	1.71–2.29
Saccharomyces cerevisiae	11	Eukaryote	355	189	1.87	1.68–2.32
Saccharomyces cerevisiae	12	Eukaryote	571	313	1.82	1.75–2.25
Saccharomyces cerevisiae	13	Eukaryote	513	273	1.88	1.74–2.26
Saccharomyces cerevisiae	14	Eukaryote	438	225	1.94	1.71–2.29
Saccharomyces cerevisiae	15	Eukaryote	597	319	1.87	1.75–2.25
Saccharomyces cerevisiae	16	Eukaryote	519	289	1.79	1.74–2.25
Caenorhabditis elegans	1	Eukaryote	2411	1140	2.11	1.88–2.12
Caenorhabditis elegans	2	Eukaryote	2962	1384	2.14	1.89–2.11
Caenorhabditis elegans	3	Eukaryote	1788	834	2.14	1.85–2.14
Caenorhabditis elegans	4	Eukaryote	2453	1149	2.13	1.87–2.12
Caenorhabditis elegans	5	Eukaryote	4325	1961	2.20	1.91–2.09
Caenorhabditis elegans	X	Eukaryote	2643	1208	2.18	1.88–2.11
	What are called Scaffolds					
Drosophila melanogaster	1	Eukaryote	1119	594	1.88	1.82–2.17
Drosophila melanogaster	2	Eukaryote	672	378	1.77	1.76–2.23
Drosophila melanogaster	3	Eukaryote	596	297	2.00	1.75–2.24
Drosophila melanogaster	4	Eukaryote	1034	555	1.86	1.81–2.18
Drosophila melanogaster	5	Eukaryote	1699	910	1.86	1.85–2.14
Drosophila melanogaster	6	Eukaryote	116	66	1.75	1.44–2.55
Drosophila melanogaster	7	Eukaryote	28	15	1.86	0.86–3.13
Drosophila melanogaster	8	Eukaryote	568	313	1.81	1.74–2.25
Drosophila melanogaster	9	Eukaryote	286	174	1.64	1.64–2.35
Drosophila melanogaster	10	Eukaryote	264	148	1.78	1.63–2.36
Drosophila melanogaster	11	Eukaryote	1756	952	1.84	1.85–2.14
Drosophila melanogaster	12	Eukaryote	142	80	1.77	1.50–2.50
Drosophila melanogaster	13	Eukaryote	2953	1593	1.86	1.90–2.11
Drosophila melanogaster	14	Eukaryote	48	24	2.00	1.14–2.86
Drosophila melanogaster	15	Eukaryote	86	44	1.95	1.35–2.65
Drosophila melanogaster	16	Eukaryote	63	28	2.25	1.24–2.76
Drosophila melanogaster	17	Eukaryote	1787	978	1.83	1.85–2.14
Drosophila melanogaster	18	Eukaryote	53	28	1.89	1.18–2.82
Drosophila melanogaster	19	Eukaryote	30	17	1.76	0.90–3.10

Table 1. Genome Data.

flip and see if such a statistical model can explain the observed chromosomal organization. It assumes that the transcription direction for each gene has a probability of 0.5 and that the probabilities are independent of each other. This flip-of-the-coin model would predict that the average number of genes on a string should lie near 2 per string. That is not what is observed in chromosomal organization. In Table 1 we list the number of genes per string (n/s) and then the 3 standard deviation range for each entry. When we compare the string lengths observed with the expected range for the coin-flip model, we see significant deviations from that in Table 1 among the archaea and bacteria. If the string lengths are consistent with this model, the n/s value has approximately a 99% chance of falling between the 3-sigma values. The clear conclusion is that this random model totally fails to account for the bacterial data. However, the gene strings of the "higher" eukaryotes, like yeast and *C. elegans*, are clearly more randomized when measured against this model. Indeed, they seem to lie within the statistically predicted values given by this flip-of-the-coin model, or nearly so. This is a surprising result when the above statements are taken into account. Moreover, a slight modification of this flip-of-the-coin model yields a very compelling fit for each of the three eukaryotes, a fit that is good enough to pass a demanding goodness-of-fit test. The standard flip-of-the-coin model assigns p = 0.5 as the probability that a given

space between adjacent chromosomes is a transition point. The modification for *D. melanogaster* is to assign p = 0.541 (the very same value of p for all 19 scaffolds!). Comparable, excellent fits are possible for yeast, with p = 0.530 (the same value of p for all 16 chromosomes!), and for *C. elegans*, with p = 0.463 (the same value of p for all six chromosomes!). The quality of the fits under p = 0.5 are not good enough to pass the demanding statistical test referred to above. We will return to the philosophical implications of this below.

Does this mean that the bacteria are incompatible with random chance when their gene strings are examined? Two other models were tested attempting to determine if a stochastic model could reasonably match the observed gene string length.

Model 2 starts with the observed data for each species—the number of genes and the number of strings. Labeling the space between two strings as a transition point, the number of transition points is equal to the number of strings (except when there is only one string—and no transition point). Model 2 next distributes the transition points randomly among the n available spaces between the genes. Then the distribution of the resulting string lengths is compared with the actual string lengths of the chromosome. In turn, the distribution of string lengths
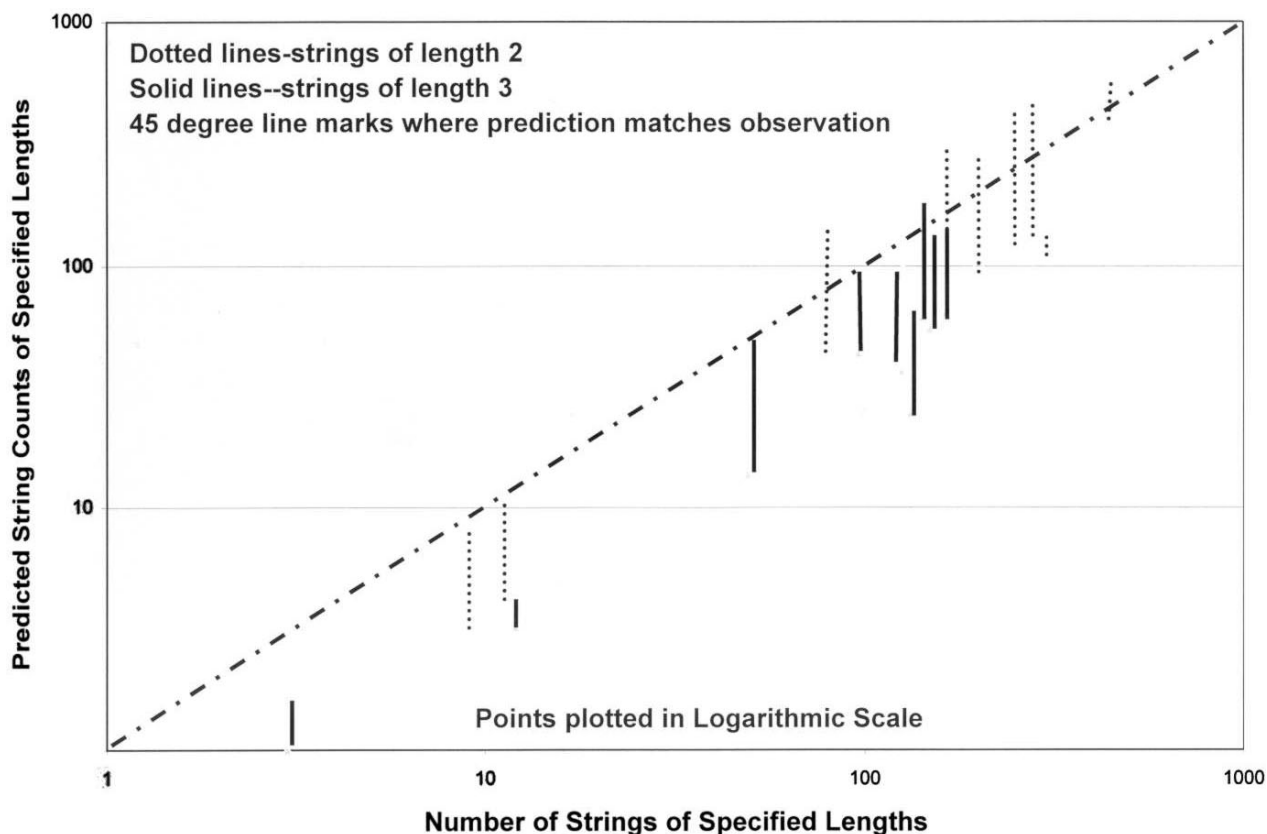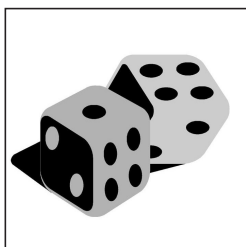

Figure 2. The results of Model 3.

Random Worms: Evidence of Random and Nonrandom Processes in the Chromosomal Structure of Archaea, Bacteria and Eukaryotes

Bacteria appear to have a genetic structure that is decidedly nonrandom. And one might be tempted to conclude that life indeed is a nonrandomly organized phenomenon. … But this does not explain why the more complex eukaryotes studied here fit Model 1 (the flip-of-the-coin model) so well, …

is compared with that in the genome. This could be called the slice-and-dice model where the sequence of genes is randomly cut and every other segment inverted. This model underestimates the actual number of strings of length 1 a bit, estimates the strings of length 2 approximately correctly, and overestimates the number of strings of length 3 a bit, and fails to account for the extraordinary length of the longest strings found in the chromosomes. There is more variability in the string lengths of bacteria than can be accounted for by randomness described by Model 2.

Model 3 assumes that the orientations of the genes (and hence the length of each gene string) correspond to the signs of n correlated (mean zero, variance one) Gaussian random variables. Placing reasonable mathematical constraints on the correlations leads to a set of three equations with four unknowns. The values of these unknowns precisely determine the distribution of the string lengths. One can solve these equations numerically to obtain a range of solutions, and, thus, a set of possible string-length distributions. It is observed in Figure 10 of Hutchison, et al. and Figure 2 in this article that these fit the observed data well.

Implications of this Work

The most obvious implication is that bacteria appear to have a genetic structure that is decidedly nonrandom. And one might be tempted to conclude that life indeed is a nonrandomly organized phenomenon. This certainly fits the preconceptions of many Christian apologists and laity who see randomness as a threat to a particular theological position. Many apologists claim that order found in biology is evidence of God's design.[14] But this does not explain why the more complex eukaryotes studied here fit Model 1 (the flip-of-the-coin model) so well, and a slight modification extremely well. This goes against the common claim by apologists that "the information required for large-scale evolution cannot come from random variations"[15] and "… the integrated complexity of developmental programs [cannot] plausibly be attributed to chance."[16]

Why is it that *C. elegans*, a nematode that is more complex than bacteria, possesses a genetic structure that appears more random-

ized than the much simpler bacteria and archaea? *C. elegans* has a nervous system, muscle cells, a hypodermis, an excretory system, and a specialized reproductive system. As we saw above, the bacteria have the genes for entire biochemical systems located in the same string; from the yeast to the *C. elegans* and on up to humans (whose incomplete and early genome information has been examined by the authors), there is no such order (imposed by need), and the gene's transcription direction is free to be mutated randomly. It is hard to avoid the conclusion that the more complex organism is more "randomized."

The data we present here has implications to the common belief that random mutations and selection cannot improve a system. Such views are illustrated by that of Davis and Kenyon who said:

A mutation in a coding gene, then, can be looked at as a random change in functional information. As a unit of functional information in the cell, a coding gene is much like a word (a unit of meaningful information) in a book. What do you think would happen if we randomly changed the letters in some of the words in this book? Would the book be improved? On the contrary, it is probable that random changes in the words of this book would decrease rather than increase the meaningful information they carry.[17]

The mathematically more randomized state of the eukaryote gene strings we examined is not quite the equivalent of randomly changing the letters in this article, but it is the equivalent of randomly reversing the direction in which the words are read. While the reading of the sentence, "Information is hurt by random reversal of the word directions" is made more difficult by laying it out as "noitamrofni si truh yb modnar lasrever fo eht drow snoitcerid" the same procedure applied to genes and gene systems seems to make for more complex beings—yeast, nematodes, fruit flies, and humans. This clearly shows that one ought not draw the analogy between words and genetic systems too closely. Words are not genetic systems and genetic systems are not words.

For those who truly understand information theory, the above results should not be

surprising. Hubert Yockey has stated that it is mathematically fundamentally impossible to tell the difference between a random sequence and one that codes for an organism![18] Indeed, the higher the information content of a sequence the more random the sequence will appear from a mathematical point of view. This is what we see with the gene strings of the higher organisms.

The ability of the modified coin-flip model to fit the observed gene string patterns seen in living systems calls into question the efficacy of William Dembski's methodology for determining design. Dembski wrote:

> Even so, complexity (or improbability) isn't enough to eliminate chance and establish design. If I flip a coin 1,000 times, I'll participate in a highly complex (i.e., highly improbable) event. Indeed, the sequence I end up flipping will be one in a trillion trillion trillion …, where the ellipsis needs 22 more "trillions." This sequence of coin tosses won't, however, trigger a design inference. Though complex, this sequence won't exhibit a suitable pattern. Contrast this with the previous sequence representing prime numbers from 2 to 101. Not only is this sequence complex, but it also embodies a suitable pattern. The SETI researcher who in the movie *Contact* discovered this sequence put it this way: "This isn't noise, this has structure."[18]

Dembski, like us, wants to conclude that living systems are designed. But Dembski's methodology for determining design, by his own admission, excludes anything that has the structure of a coin-flip model, which is exactly what we find in the patterns of the gene strings. Are we to conclude from this that the cells are not designed? No! What we can conclude is that Dembski's model is inadequate to the task he intends. Once again, as Yockey notes, organized sequences will appear the same as randomly generated sequences. As an example, consider the sequence "yjrvsyonstrfdjoty" which is a Caesar substitution cipher for "thecatintheredshirt." Both sequences contain the same meaning and semantical information. The only reason the latter, decrypted sequence does not look randomized is because of years of training in how to read letters in that order.

Conclusions

We have presented evidence for both random and non-random features of the chromosomal structure. The asymmetries in the gene directionality and the bacterial deviations from the predictions of our models must be due to nonrandom forces and mechanisms required for life to exist. But at the same time, there is already much evidence that the chromosomal organization of the eukaryotes is nearly consistent with a simple flip-of-the-coin model and show more randomization than do the bacteria and archaea.

It must be concluded from our study that many of the apologetical statements about living systems do not find unequivocal empirical support. The observation that the chromosomes of more advanced animals appear more random does not support the concept that information and randomness are incompatible. The higher informational content of the higher organisms will most assuredly be measured as greater randomness.


Many of the apologetical statements about living systems do not find unequivocal empirical support.

Secondly, the failure to find a completely suitable stochastic model for the bacteria and archaea does not mean that randomness does not operate on their chromosomal arrangement. What it means is that there are functional limitations caused by the differences in how bacteria and archaea, on the one hand, and eukaryotes, on the other, process the DNA's information into protein. Those differences cause bacteria to appear less random, but this also means that they appear more ordered. And since order is a property of static things (like highly ordered, but totally inert, mineral crystals), more order in a living system is an indication of less complexity.

Methods for determining design need to be robust enough to conclude that the structure of a cell is designed. If it cannot, then one of two propositions therefore must be true: (1) either the cell is not designed; or (2) the methodology is flawed. We prefer to believe that the current Dembski methodology is flawed.

The preference for short gene strings, which our data demonstrates, may have implications for origin of life issues. It shows that life at this level of structure is built up of repetitive, simpler systems which are amalgamated into the larger unit. Contrary to the perception that higher organisms should use larger and more complex organizational systems for their genomic information, they actually use a system of gene strings which are small and simple rather than large and complex.

Finally, Christian apologists need to incorporate chance and randomness into their world views. It is clear that the Bible teaches that God controlled chance and randomness at several crucial junctures in history. If he did this, then the controlling of chance and randomness in biology should be equally possible. Apologists should not ignore the observational fact that as we go from simple creatures to the more complex organisms, chromosomal organization appears to be much more influenced by random processes. Indeed, chromosomal structure at the complex end
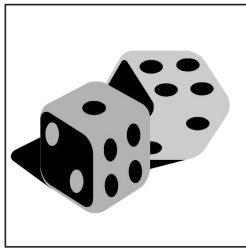
Random Worms: Evidence of Random and Nonrandom Processes in the Chromosomal Structure of Archaea, Bacteria and Eukaryotes

of the spectrum matches the predictions of very simple stochastic models. It does no good to claim that chance or random processes cannot produce more complex organisms, when those very organisms are measurably more random than are the simpler creatures.

In conclusion, we should all remember what God said to Job: "'Will the one who contends with the Almighty correct him? Let him who accuses God answer him!' Then Job answered the LORD: 'I am unworthy—how can I reply to you? I put my hand over my mouth. I spoke once, but I have no answer—twice, but I will say no more'" (40:2, NIV).

It seems to us that God is not scolding Job for his ignorance but rather for his lack of humility—by attempting to explain how God conducts his activities. Applied to our present discussion, we would dare to suggest that a significant degree of humility is needed by all Christians, to accept—in agreement with various biblical teachings—that God is perfectly capable of fully designing the various life forms we observe, however he pleases, in a way we surely do not understand, while presenting his handiwork to us mere mortals in a format that shows strong evidence of randomness. ✺

> *It does no good to claim that chance or random processes cannot produce more complex organisms, when those very organisms are measurably more random than are the simpler creatures.*

Notes

1. Henry M. Morris, "The Compromise Road," *Impact* 177 (March 1988): I, ii.
2. Lee Spetner, *Not by Chance* (Brooklyn: The Judaica Press, Inc., 1998), vii.
3. Jonathan Wells, "Making Sense of Biology," *Touchstone* (July/August 1999): 55.
4. Adnan Oktar, "The Evolution Deceit," www.hyahya.org/16understanding03.php, accessed Sept. 23, 2001.
5. William A. Dembski, "Signs of Intelligence," *Touchstone* (July/August 1999): 80.
6. ——, "On the Very Possibility of Intelligent Design," in J. P. Moreland, ed., *The Creation Hypothesis* (Downer's Grove: InterVarsity Press, 1994), 116; A. E. Wilder-Smith, *The Natural Sciences Know Nothing of Evolution* (San Diego: Master Books, 1981), 85; Phillip Johnson, "Creator or Blind Watchmaker," *First Things* (Jan. 1993): 12; Phillip Johnson, "Immodest Ambitions," *Books & Culture* (Sept. Oct. 1995): 29; Michael Behe, *Darwin's Black Box* (New York: The Free Press, 1996), 191–2; Robert F. DeHaan and John L. Wiester, "The Cambrian Explosion," *Touchstone* (July/August 1999): 65.
7. See Exod. 28:30; Lev. 8:8; Deut. 33:8; Ezra 2:63; and Neh. 7:65. See also "Urim and Thummim" Trent C. Butler, ed. *Holman Bible Dictionary* (Nashville: Holman Bible Publishers, 1991) for more on this.
8. Dr. Robert Gange, *Origins and Destiny* (Waco, TX: Word, 1986), 73.
9. Available in the English translation of *Foundations of Probability Theory* (New York: 1950).
10. Gregor Mendel *Experiments in Plant Hybridization*, (1865).
11. www.ncbi.nlm.nih.gov/PMGifs/Genomes/allorg.html
12. Most techniques used are found in any statistical textbook. Some special techniques of note are: F. N. David, "A Note on the Evaluation of the Multivariate Normal Integral," *Biometrika* 40 (1958): 458–9; W. F. Sheppard, "On the Application of the Theory of Errors to Cases of Normal Distribution and Normal Correlation," *Philosophical Transactions of the Royal Society of London Series A* 192 (1899): 101–67; W. F. Sheppard, "On the Calculation of the Double Integral Expressing Normal Correlation," *Transactions of the Cambridge Philosophical Society* 19 (1900): 23–66.
13. Clyde Hutchison, Michael Montague, Glenn Morton, Michelle Opp, Gabor Pataki, and Gordon Simons (2001, unpublished manuscript); see also Gordon Simons and Glenn Morton, "The Gene-Orientation Structure of Eukaryotes," *Journal of Theoretical Biology* 22, no. 4 (2003): 471–5.
14. Henry M. Morris, "The Power of Energy," in Walter E. Lammerts, ed., *Scientific Studies in Special Creation* (Grand Rapids: Baker Book House, 1971), 66; Randy L. Wysong, *The Creation-Evolution Controversy* (Midland, MI: Inquiry Press, 1976), 257; Henry Morris, *The Remarkable Birth of Planet Earth*, (Bloomington, MN: Bethany House, 1972), 1; ——, *Creation and the Modern Christian* (El Cajon, CA: Master Book Publishers, 1985), 214–5; A. E. Wilder-Smith, *Man's Origin, Man's Destiny* (Wheaton, IL: Harold Shaw, 1968), 57.
15. Lee Spetner, *Not by Chance* (Brooklyn: The Judaica Press, Inc., 1998), vii.
16. Wells, "Making Sense of Biology," 55; see also Rick Wade, "Defeating Darwinism," in Ray Bohlin, ed., *Creation, Evolution, & Modern Science* (Grand Rapids, MI: Kregel Publications, 2000), 98.
17. Percival Davis and Dean H. Kenyon, *Of Pandas and People* (Dallas: Haughton Publishing Co., 1993), 66.
18. Hubert Yockey, *Information Theory and Molecular Biology* (Cambridge: Cambridge University Press, 1992), 81–2.
19. Dembski, "Signs of Intelligence," 79.

X4