

Can biologically active sequences come from random DNA?

Robert W. Carter

A recent report that random DNA sequences can be a source of biological novelty is being used to support evolution. The authors concluded that biologically important novelty was trivial to generate. However, they drew multiple premature conclusions from their work, and they made no attempt to correlate their sequences with known biological function. In this follow-up study, the standard sequence comparison tool BLASTn was used to probe for similarities between their random sequences and the *E. coli* genome. In most cases, a 20–40-bp section was identified that had a high degree of similarity (up to 100%) to a small portion of a known *E. coli* gene. In the majority of cases, the random DNA ran in the reverse direction from that of the gene. This strongly indicates that a specific subsection of the RNA transcript, and not the protein product of the randomized DNA, was the active agent. This size range resembles that of many biologically active RNA molecules, specifically microRNAs, that are known to have a major influence in regulating expression of many different genes. There is no evidence here that random DNA supports evolutionary theory. Instead, random RNAs inserted into the cell help us learn about the amazing complexity of genetic regulation.

Recently, *Biologos*¹ fellow Dennis Venema reiterated the common evolutionary claim that new biological functions can easily arise from random mutation.² As his first example, he used the *nylonase* gene. For several decades, evolutionists have been claiming the existence of the *nylonase* gene as *prima facie* evidence for evolution. The fact that a bacterium was able to ‘evolve’ the ability to digest a man-made polymer in just a few years was seen as a triumph of evolutionary predictions. But the early claims that attempted to describe how the gene arose fell short of reality. Instead of a ‘frame shift’ in a gene that caused the new ability to arise, an enzyme that already had the ability to digest similar molecules was fine-tuned by the bacterium to break the nylon bond. But this was done in a copy of the original gene on a plasmid. The original was left untouched.³ Since some bacteria already had the ability to degrade a similar bond (the amide bond found in all proteins), and since the enzyme already had a limited ability to degrade nylon, it only took a few minor changes in the backup copy of the enzyme to allow for more efficient nylon degradation. Thus, the *nylonase* gene is much better suited to supporting design arguments than to supporting evolution in general.

However, Venema brought up a second example, which comes from new research, where the experimenters supposedly found a high frequency of biologically active properties in random DNA sequences. An analysis of this new study will be the focus of this paper. But Venema shows his bias by asking, “Just how easy is it to obtain a functional gene from random DNA sequence? And consequently how likely is it that *de novo* gene origination is a common occurrence?” In both sentences he uses the term ‘gene’ without grappling

with the nuances of the modern concept of genes and genetic information. Is it true that a random sequence, when inserted into a cell, has the capacity to take on the role of a ‘gene’?

The authors of the study under question, Neme *et al.*, state, “Intriguingly, the highest rates of *de novo* emergence are always found in the evolutionarily youngest lineages.”⁴ This defies evolution, for it would mean that evolutionary rates are speeding up over time. Using circular logic, they are claiming that more ancient sequences evolve more slowly because they are more conserved.⁵ This does nothing to help their argument that new function can arise easily from random DNA and illustrates how our opponents often play fast and loose with important concepts and definitions.

In their study, Neme *et al.* generated millions of random 150-bp DNA sequences and inserted them into a bacterial plasmid. They then induced *E. coli* to absorb these plasmids. The plasmid carries an ampicillin resistance gene so any non-transformed bacteria would die when grown in the presence of the antibiotic. It also carries an inducible promoter that would turn on transcription of the random DNA sequence when exposed to IPTG.⁶ The plasmid also carries a built-in stop codon. This guarantees that a protein with a randomized centre comprising 50 amino acids would be made after the gene was transcribed. This is about the size of a typical protein domain, but note that evolution must explain how entire proteins evolve, not just disconnected subsections of proteins. Also, three of the 64 codons are stop codons; thus, stops should occur every 21.3 bases on average. Therefore, most of their sequences would not have been expected to produce a full-length protein.

When grown in mixed culture, they were surprised to discover many clones, the growth rates of which were affected by the presence of the random DNA. Although most of the random DNA sequences they scored caused a decrease in growth rate, some did the opposite. They took this to indicate that some of the random sequences affected the cells enough that selection (either purifying or positive) could have acted upon them.

The experiment is ingenious, and, as an intellectual exercise, reveals intriguing lines for future enquiry. Technically, they did nothing wrong. However, they made several critical errors when attempting to extract evolutionary connotations.

Their first error was one of applicability. We know that nothing in life produces truly random sequences, and no part of evolutionary theory (after the origin of life) starts with randomized nucleotides. The typical protein consists of multiple interspersed functional domains and disordered regions.⁷ This does not mean the intrinsically disordered regions (IDRs) have no function, however; they are involved in multiple important cellular processes from affecting protein folding to influencing protein assembly. IDRs also have distinct compositional biases (i.e. they have more charged and polar amino acids and fewer amino acids with bulky hydrophobic groups). They are not truly ‘random’ (see previous reference for a detailed discussion) and should not serve as a source of truly random DNA for evolutionary purposes. Unlike humans and higher organisms, bacteria have little ‘junk DNA’,⁸ so this cannot be the source of new functional novelty.

Second, the authors failed to address how much time would be required to sample these random sequences in real life. Sanford *et al.* studied how long it would take a random functional string to appear in a human-like population.⁹ Their model results indicate that it would take approximately 84 million years for random mutation to produce, and for natural selection to fix, even a strongly favoured 2-nucleotide string. It would take more time than the history of life on Earth to fix a 6-nucleotide string.¹⁰ In a similar vein, O’Micks studied the evolution of bacterial gene promoters via random mutation and concluded it was virtually impossible.¹¹

This ‘waiting time problem’ is a significant hurdle for evolution to cross. Bacteria like *E. coli* have much shorter generation times and much higher population sizes than humans, and so might be able to experiment with much more random DNA over time. Yet, Neme *et al.* made no estimate concerning how much time this might take, even allowing for the sudden appearance of 150-bp random sequences that can be transcribed and translated in the cell.

Third, the sequence space they explored was probably orders of magnitude greater than what life could ever experience. There are four nucleotides in DNA, thus the

potential for 4^{150} ($>2 \times 10^{90}$) theoretical sequences 150 nucleotides in length. Since they were dealing with μg quantities of DNA, they did not even begin to exhaust the possibilities. However, they did test tens of millions of different sequences.¹² Also, most genes do not have to be perfect to manufacture either a functional RNA or protein. Thus, they may have sampled a much greater proportion of protein or functional RNA space than one might assume at first.

To develop these thoughts further, another standard laboratory procedure needed to be applied to their sequence data, one which is available to them, yet they curiously failed to perform: BLAST.

Methods

In their supplementary information, Neme *et al.* provided a list of 713 random 150-bp sequences (and the 50-amino-acid translated proteins) they determined were biologically active. They also flagged each sequence ‘up’ or ‘down’ to indicate whether it would have a positive (+) or negative (–) effect on bacterial numbers over time. They cloned the random sequences into a specific plasmid vector, leaving a DNA sequence with this formula:

```
ATGAAGCTTAGC...N150...GCATTGGTCGACTACAAGGA
CGATGACGACAAGTGA
```

where N_{150} represents the 150-bp randomized DNA sequence. This translates into a protein with this formula:

```
MetLysLeuSer...AA50...AlaLeuValAspTyrLysA
spAspAspAspLysSTOP
```

where AA_{50} represents the randomized string of 50 amino acids.

In their paper, they reported analyses on a small subset of the active sequences. Specifically, they tested the activity of clones 3 (+), 8 (+), 53 (–) and 119 (–). They also assayed clones 4 (+), 32 (+), and 600 (+) in competition experiments. They did not include clone 600 in the sequence list, for unexplained reasons. Clone 605 was used here instead, since they listed it as ‘similar to 600’.

The >700 biologically active clones Neme *et al.* listed should not have been in any particular order, so the first 10 ‘up’-regulating and the first ten ‘down’-regulating clones were treated as a representative sample. I also examined all seven of the clones they specifically assayed in competition experiments. I searched for similar sequences among these 27 clones using the standard BLASTn tool (v. 2.6.1).¹³ There are many different parameter settings that affect BLAST results, but, knowing that they used short sequences with potentially little similarity to living things, and after some experimentation, I set the Expect Threshold to 20 (higher than normal) and the Word Size to 11 (smaller than normal) to account for these difficulties. At low word sizes, the

trailing FLAG sequence received many hits due to the popular use of this vector in many different studies, so the leading and trailing plasmid vector sequences were trimmed prior to any reported BLAST search. I used BLAST directly on the *E. coli* genome first. To broaden the applicability of these results, I also used BLAST against a set of curated diverse genomes (refseq_representative_genomes). I also used the random sequence generator at bioinformatics.org¹⁴ to create multiple random nucleotide strings 150 to 1,500 long. This was done to create a set of random sequences that were not first filtered for activity in *E. coli*. After a few initial trials, I opted to not search the entire NCBI nucleotide collection (with the exception of the longest random string) because this generates many non-biological, engineered, and duplicate hits. The purpose was not to identify every biological sequence that matched these random sequences, but only to identify and characterize a few high-scoring matches, if they existed.

Results

Neme *et al.* claimed their random sequences were synthesized as “equimolar mixes of A, C, G, and T at every position”, but we do not know if they validated this. The 713 biologically active sequences they reported had decidedly non-random nucleotide frequencies (figure 1). An even distribution would mean all nucleotides should have a frequency of 0.25, but the reported sequences were rich in G (0.33 +/- 0.03 SD) and depauperate in A (0.18 +/- 0.03 SD). The other two nucleotides were exactly at expectation (0.25 +/- 0.04). They did not perform this simple measure and may have noticed something was amiss if they had. Instead of ‘random’ sequences showing functionality, the ‘biologically active’ sequences had highly skewed nucleotide ratios, indicating that something decidedly non-random was occurring with the *E. coli* populations that carried these sequences.

They did not analyze the nucleotide composition of their clones, but they did perform an analysis on amino acid frequencies. Since one of their (and Venema’s) assumptions was that the synthesized proteins would be the active agents in their assay, they incorrectly state that the amino acid composition provides “potentially more information than nucleotide composition of the underlying RNAs”. They found no significant differences from random expectations, but they did note that specific amino acids were less common (E, I, N, Q, and T) or more common (C, D, G, R, and S) in the random sequences than in *E. coli*. This pattern does not match that found in IDRs (see Discussion). After adjusting for codon frequency,¹⁵ I calculated the nucleotide frequency within the

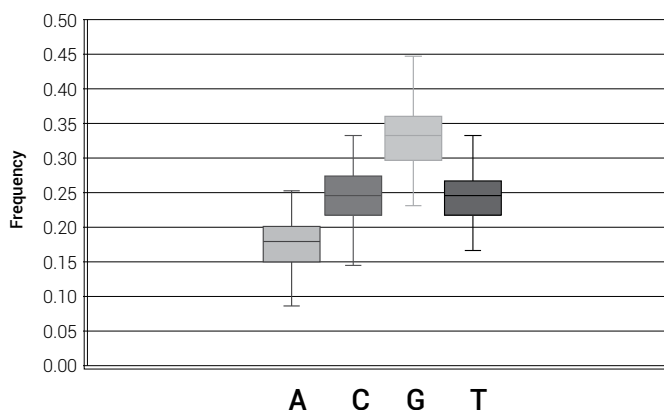


Figure 1. Box-and-whisker plot of the nucleotide frequencies within the 713 biologically active random sequences reported by Neme *et al.* The mean is represented by the horizontal line within the box. The data are divided into quartiles, represented by the top and bottom edges of the boxes and the top and bottom ends of the whiskers.

64 codons used in *E. coli*. I then calculated the nucleotide frequency of the codons for the amino acids that were more and less common than expected. The results were an exceedingly close match to that of the nucleotide composition within the clones. That is, the codons for the amino acids that appeared at higher-than-expected frequencies had less A and more G than average, and vice versa (table 1). Thus, the amino acid composition in the putative protein products was a simple function of the uneven nucleotide composition in the random sequences. This is evidence that the random sequences are acting on the RNA/DNA level.

The very first BLAST search produced a startling result: clone 2 contains a 27-bp subsection of the *E. coli* sensor histidine kinase gene (figure 2). This gene happens to be involved in citrate metabolism.

The text output of a search includes information on the organism and/or strain name, where the match occurs along the search and target string, and in which nucleotides are identical. In this case, 24 of the 27 nucleotides (89%) are identical between the two (figure 3):

Table 1. The expected usage of each of the four nucleotides in the proteins coded in the biologically active random DNA sequences, after adjusting for codon usage. The first row shows the frequency of each nucleotide among all codons in *E. coli*. The second row shows the frequencies among the amino acids specifically flagged as less abundant (R, D, C, S, and V) in *E. coli* compared to the random sequences. The third row shows the frequencies among the amino acids specifically flagged as more abundant (N, E, Q, I, and T).

table a	f(A)	f(C)	f(G)	f(T)
All codons in <i>E. coli</i>	24.2%	25.5%	25.9%	24.4%
Less common in <i>E. coli</i>	12.8%	14.8%	45.2%	27.2%
More common in <i>E. coli</i>	47.3%	27.0%	10.4%	15.3%

Table 2. BLAST results for multiple clones compared to *E. coli*. Included here are the first 10 'up' and the first 10 'down' strains, and each strain assayed by Neme *et al.* in competition experiments. The matching gene name is dependent on the annotation data provided by the contributor of the sequence, thus not all annotations are of the same quality. 'Direction' indicates whether the clone runs in the same direction as the gene in question. Due to the potential for frame-shifting, only 1/3 of the clones that match in the forward direction are expected to produce a protein that matches the relevant section of the gene in question.

Clone	+/-	Match (bp)	Match (%)	Matching gene name	Direction
2	+	27	89	sensor histidine kinase	Reverse
3	+	35	83	HK97 family phage portal protein	Reverse
4	+	18	100	plasmid stability factor protein StbB	Reverse
5	+	38	80	peptide permease	Reverse
6	+	36	83	plasmid transposase	Reverse
		21	95	'conserved hypothetical protein'	Forward
7	+	28	86	glycine/betaine ABC transporter	Reverse
8	+	21	95	control region, just upstream of an iron ABC transporter gene	Forward
9	+	23	91	translation initiation factor Sui1	Reverse
				glucuronide uptake porin UidC translation initiation factor	Reverse
				ligand-gated channel protein translation initiation factor	Reverse
				stress response translation initiation inhibitor YciH	Forward
10	+	22	95	vitamin B12 import system permease BtuC	Forward
11	+	18	100	scaffolding protein	Forward
12	+	27	0.89	gene 'AHM28362' (strain ST540, position 476,362)	Reverse
				gene 'AHM35289' (strain ST540, position 3,049,395)	Reverse
				gene 'AHM39887' (strain ST540, position 3,149,293)	Reverse
13	+	22	95	dyhydroneopterin aldolase	Reverse
14	+	-	-	(no significant similarity found)	n.a.
32	+	26	88	L-galactonate transporter	Forward
		21	95	Nickel-responsive regulator	Forward
		23	91	a tail tape-measure protein	Reverse
600	+	43	79	integrase (aka trigger factor)	Forward
30	-	-	-	(no significant similarity found)	n.a.
35	-	21	95	plasmid DNA, 200 bp above a known gene	n.a.
50	-	-	-	(no significant similarity found)	n.a.
51	-	24	92	putative prophage protein	Forward
53	-	-	-	(no significant similarity found)	n.a.
57	-	29	86	glutamate-tRNA ligase	Forward
59	-	-	-	(no significant similarity found)	n.a.
60	-	27	89	<i>E. coli</i> pathogenicity island 1	n.a.
61	-	24	92	plasmid DNA, 150 bp downstream and 200 bp upstream of unnamed genes	n.a.
		25	92	unnamed gene	Forward
65	-	-	-	(no significant similarity found)	n.a.
70	-	18	100	hypothetical protein	Reverse
119	-	18	100	tyrosine protein kinase	Forward

The statistics pertaining to this situation seem perplexing at first. On the one hand, a 15-nucleotide sequence would be expected to be found once in a billion random nucleotides, and a 30-nucleotide sequence once in every 10¹⁸ random nucleotides. These numbers are much larger than the *E. coli*

genome (of approximately 4.6 million bases). But there are several mitigating factors that greatly increase the probability of a significant hit.

First, the matching sequences do not have to be exact. There are many permutations of a 15-bp nucleotide string with one or more allowed ambiguous bases in random positions along that string.

Second, one major mistake the authors made was to assume that DNA is random. It is not. Certain combinations of letters are favoured, and others disfavoured, at all levels of organization. Unlike the DNA of higher organisms, the four nucleotides in *E. coli* are found at approximately the same frequency (24.6–25.4%). However, this is not true of the 16 dimers (4.6–8.3%), and the spread increases with increasing word size (figure 6). In fact, departures from random expectations can be found among any set of n-mers, even after accounting for the frequencies of the smaller n-mers. Thus, even though there is an astronomical number of nucleotides 150-bp in length, due to the non-random nature of biological DNA a certain subset of those combinations are highly likely to match significant portions of DNA.

Failure to take into account the non-randomness of biological DNA at all levels led a team of computer scientists at IBM to mistakenly identify millions of ‘pyknons’ in the human genome.¹⁸ These seemed like a ‘code within a genetic code’, and would have been an exciting discovery.¹⁹ However, they merely found repeating subunits of the already-known and well-characterized Alu elements that happened to permeate the genome.

Neme *et al.* made additional errors when saying things like, “Contrary to expectations, we find that random sequences with bioactivity are not rare.” This is patently untrue. They discovered approximately 700 active sequences. Out of the millions of sequences they started with, this represents a very small percentage

Table 3. Significant BLAST hits from a curated list of representative genomes for the seven assay clones

Clone	# Hits	Identity range	Species
3 (+)	3	83–93%	<i>Castor canadensis</i> [beaver] <i>Nocardiopsis ganjiahuensis</i> [soil bacterium] <i>Mesorhizobium sp.</i> [Gram-negative soil bacterium]
4 (+)	4	80–88%	<i>Callithrix jacchus</i> [common marmoset] <i>Parus major</i> [great tit] <i>Pseudopodoces humilis</i> [ground tit] <i>Rubrobacter xylanophilus</i> [Gram-positive bacterium]
8 (+)	2	86–96%	<i>Ixodes scapularis</i> [deer tick] <i>Fusarium graminearum</i> [wheat head blight fungus]
32 (+)	2	91%	<i>Cladophialophora immunda</i> [black yeast] <i>Leisingera aquaemixtae</i> [Gram-negative bacterium]
53 (-)	5	83–100%	<i>Microbacterium sp.</i> [Gram-positive bacillus] <i>Hipposideros armiger</i> [great roundleaf bat] <i>Branchiostoma floridae</i> [lancelet] <i>Pogona vitticeps</i> [central bearded dragon] <i>Castor canadensis</i> [beaver]
119 (-)	3	82–90%	<i>Corynebacterium xerosis</i> [Gram-positive bacterium] <i>Pseudonocardia dioxanivorans</i> [bacterium] <i>Pelagibaca bermudensis</i> [marine roseobacterium]
600 (+)	2	91–94%	<i>Actibacterium mucosum</i> [marine alphaproteobacterium] <i>Cladophialophora immunda</i> [black yeast]

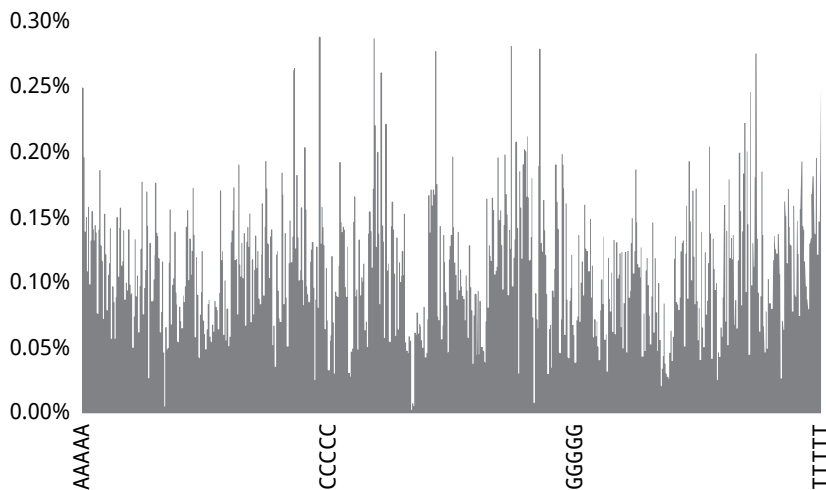


Figure 6. The frequencies of the 1,024 pentamers in the *E. coli* genome (strain REL606) are far from random. They range from 0.0025% to 0.29%. With a genome size of 4,629,812 nucleotides, there is more than enough data to generate a robust average of each frequency, so the data presented are not sampling artifacts. The range of frequencies only increases with increased word size.

of all sequences assayed (literally ‘one in a million’). While we have no idea how many of these random sequences were severely detrimental to the cell because these would quickly disappear from the culture, one would expect that most random sequences would have no effect at all.

They make an additional error by assuming that the random sequences add biological novelty to the cell. There is, in fact, no evidence for this. The majority of sequences I analyzed had a highly significant match to a known gene or what might be assumed to be a control region of a known gene. If this were not the case, one might be able to argue that short, random *proteins* can create biological novelty. Instead, it appears that short, random *nucleotides* interfere with cellular operations.

The high proportion of sequences that match the reverse compliment of a known gene demonstrate that orientation is unimportant. But functional areas can include non-genic areas like promotor regions. Thus, the protein sequence, at least in most cases, though perhaps all, is also unimportant.

If these ‘bioactive’ DNA sequences are not producing functional proteins, they must be acting on the level of RNA–RNA or RNA–DNA interactions. The annealing temperatures of ribonucleic acids depend on their length and percent identity. Biological function in this case does not depend on sequence specificity. Also, the triple-hydrogen bonding G and C bind more tightly than the double-hydrogen bonding A and T, meaning sequences rich in G and C have a higher melting temperature (the temperature at which the two nucleic acids will separate in solution). The placement of G and C along the strand also impacts annealing, with terminal Gs and Cs serving to anchor the strand more so than internal ones. The skewed frequencies of A (low) and G (high) seen in the data are quite interesting in this context.

Why do we not see longer or shorter ‘bioactive’ sequences? First, due to the sheer number of permutations along a DNA strand, as the search string gets longer, the expected number of matches drops off exponentially. Second, it may be that the BLAST algorithm is cutting off less-than-perfect, but still functional, leading or trailing sequences that are beneath the detection threshold. Third,

shorter sequences will not have a high enough annealing temperature to interact directly with the genome.

What we see are the sequences at just the right length. Their RNA transcripts are long enough (20–40 nucleotides) that they could bind tightly to both RNA and DNA under physiological conditions (e.g. 37°C). The two RNA ends that have no match to the surrounding sequence would not anneal, however. This will affect the annealing of the ‘random’ RNA strand, but to an unknown extent. The RNAs produced in their experiment were on the order of 700 nucleotides, only 150 of which were the ‘random’ component. Since these

Table 4. BLAST results generated from random sequences. Tests 1–5 used 150-nucleotide test sequences. Test 6 used a 1,500-nucleotide test sequence. Genomic contexts (if available) are provided for test 6.

Test	# Hits	Identity Statistics	Species
1	3	38 bp 84%	<i>Vibrio bivalvicida</i> [pathogenic bacterium] <i>Vibrio tubiashii</i> [pathogenic bacterium] (match identical to above) <i>Aggregatibacter aphrophilus</i> [proteobacterium] (non-overlapping with above)
2	3	28–46 bp 80–93%	<i>Babesia equi</i> [protozoan] <i>Baudoinia panamericana</i> [fungus] <i>Panthera pardus</i> [leopard] (all overlapping)
3	3	23–43 bp 81–100%	<i>Numida meleagris</i> [helmeted guineafowl] <i>Schistosoma mansoni</i> [parasitic flatworm] (overlapping) <i>Helicoverpa armigera</i> [cotton bollworm] (non-overlapping)
4	5	23–49 bp 81–100%	<i>Branchiostoma floridae</i> [lancelet] <i>Asparagus officinalis</i> [asparagus] <i>Algoriphagus marincola</i> [marine sediment bacterium] <i>Paraphaeosphaeria sporulosa</i> [fungus] <i>Oncorhynchus kisutch</i> [coho salmon]
5	2	32–36 bp 89–91%	<i>Plasmodium berghei</i> [protozoan] <i>Labrus bergylta</i> [Ballan wrasse]
6	8	24–47 bp 83–100%	<i>Pseudomonas fluorescens</i> [Gram-negative bacterium] plasmid pQBR55, gene CEK42535, reverse <i>Yersinia ruckeri</i> [Gram-negative bacterium] gene ARZ027031, reverse <i>Apteryx australis mantelli</i> [North Island (NZ) brown kiwi] <i>No gene annotations provided</i> <i>Angiostrongylus costaricensis</i> [parasitic nematode] <i>No gene annotations provided</i> <i>Mus musculus</i> [house mouse] (3 identical matches) immunoglobulin heavy chain complex (Igh), Reverse <i>Mus musculus</i> [house mouse] a few hundred bp above and below two copies of MER1, a gene involved in chromosome pairing in yeast

are long compared to the oligomers flagged by BLAST, it is quite possible that they might not anneal to the bacterial DNA directly. Instead, they may operate through RNA interference, soaking up regulatory RNAs that would otherwise anneal to those 20–30 bp sections of the bacterial genome. It is also possible that they could interfere with translation by annealing to the mRNA in those short target areas.

Our understanding of the role of RNA in the cell has exploded over the previous decade. Specifically, microRNAs are short, non-coding RNAs, approximately 22-bp in size, that play multiple roles in genomic regulation.²⁰ They bind to transcribed mRNA, rendering them inactive and preventing protein translation. But short RNAs can also bind to DNA. The evidence presented in this paper suggests that Neme *et al.* stumbled upon a set of short RNA sequences that interfere with normal cellular gene regulation patterns.

Conclusion

By introducing random RNAs into the cell, Neme *et al.* inadvertently changed the genomic regulation patterns of *already existing genes*. No new functions were added. No evolution has taken place. While the experiment was ingenious, the conclusions they derived from it were unwarranted. Venema was premature in his praise.

Acknowledgments

I thank Shaun Doyle for his critical review of an earlier draft of this manuscript as well as the efforts of two anonymous reviewers.

References

1. Cosner, L., Evolutionary syncretism: a critique of BioLogos, creation.com/biologos-evolutionary-syncretism, 7 September 2010.
2. Venema, D., Biological Information and Intelligent Design: new functions are everywhere, biologos.org/blogs/dennis-venema-letters-to-the-duchess/biological-information-and-intelligent-design-new-functions-are-everywhere, 18 May 2017.
3. Batten, D., Nylon-degrading bacteria: update, creation.com/nylonase-update, 19 May 2017.
4. Neme, R. *et al.*, Random sequences are an abundant source of bioactive RNAs or peptides, *Nature Ecology and Evolution* **1**:0127, 2017.
5. Cserháti, M., Creation aspects of conserved non-coding sequences, *J. Creation* **21**(2): 101–108, 2007; creation.com/images/pdfs/tj/j21_2/j21_2_101-108.pdf.
6. IPTG stands for Isopropyl β-D-1-thiogalactopyranoside. It is a chemical mimic of allolactose that is used to induce protein expression in this system. IPTG binds to the *lac* repressor, freeing up the *lac* gene for transcription while at the same time exposing a strong promoter just upstream of the engineered sequence.
7. Van der Lee, R. *et al.*, Classification of Intrinsically Disordered Regions and Proteins, *Chem. Rev.* **114**(13):6589–6631, 2014.
8. Of course, this is a misnomer and the term would be retired were it not for evolutionary intransigence. See Carter, R.W., The slow, painful death of junk DNA, *J. Creation* **23**(3):12–13, 2009; creation.com/junk-dna-slow-death.
9. Sanford, J. *et al.*, The waiting time problem in a model hominin population, *Theoret. Biol. and Med. Modelling* **12**:18, 2015.
10. Note that this is not the ‘time to first appearance’, which is much less than the time to fixation since >99.9% of all new mutations in a hominin-like population are lost over time. See Rupe, C.L. and Sanford, J.C., Using numerical simulation to better understand fixation rates, and establishment of a new principle; in: Horsetmeyerm M. (Ed.) Haldane’s Ratchet, *Proceedings of the Seventh International Conference on Creationism*, Creation Science Fellowship, Pittsburgh, PA, 2013.
11. O’Micks, J., Promoter evolution is impossible by random mutations, *J. Creation* **30**(2): 60–66, 2016.
12. Personal communication with the corresponding author (D. Tautz) confirmed that at least ‘millions’ of clones were in their library. How many were tested is unknown, since they did not report the molar concentrations of the DNA, nor how large a quantity they used, nor the estimated transformation efficiency in the electroporation step. However, in similar experiments I would typically take 30 μl of cells at 10⁹ to 10¹⁰ cells/ml, transform with a plasmid in pg/ml to μg/ml concentration, and almost always get > 50% transformation efficiency. This would put the number of clones at least in the millions per transformation.
13. Altschul, S.F. *et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucl. Acids Res.* **25**:3389–3402, 1997.
14. bioinformatics.org/sms2/random_dna.html.
15. I used the codon frequency table at www.kazusa.or.jp/codon.
16. Carter, R.W., Splicing and dicing the human genome: Scientists begin to unravel the splicing code, creation.com/splicing-and-dicing-the-human-genome, 29 June 2010.
17. Carter, R.W., Can mutations create new information? *J. Creation* **25**(2):92–98, 2011; creation.com/mutations-new-information.
18. Rigoutsos, I. *et al.*, Short blocks from the noncoding parts of the human genome have instances within nearly all known genes and relate to biological processes, *PNAS* **103**(17):6605–6610, 2006.
19. Meynert, A. and Birney, E., Picking pyknons out of the human genome, *Cell* **125**: 836–838, 2006.
20. Arneigh, M.R., It’s a small world—microRNA cuts evolution down to size, *J. Creation* **27**(2):85–90, 2013.

Robert Carter received his Bachelor of Science in Applied Biology from the Georgia Institute of Technology in 1992 and his Ph.D. in Coral Reef Ecology from the University of Miami in 2003. He has studied the genetics of fluorescent proteins in corals and sea anemones and holds one patent on a particular fluorescent protein gene. His current research involves looking for genetic patterns in the human genome and the development of a biblical model of human genetic history. He works as a speaker and scientist at CMI-US.