ISSN: 2162-4453 Volume 10, Number 1, March 2020

Scientific Research Publishing

Open Journal of Genetics

ISSN: 2162-4453

https://www.scirp.org/journal/ojgen

Journal Editorial Board

ISSN: 2162-4453 (Print), 2162-4461 (Online)
https://www.scirp.org/journal/ojgen

Editor-in-Chief

Editorial Board

Scientific Research Publishing

Open Journal of Genetics, 2020, 10, 1-15
https://www.scirp.org/journal/ojgen
ISSN Online: 2162-4461
ISSN Print: 2162-4453

Table of Contents

Volume 10 Number 1 March 2020

Open Journal of Genetics (OJGen)
Journal Information

SUBSCRIPTIONS

The *Open Journal of Genetics* (Online at Scientific Research Publishing, https://www.scirp.org/) is published quarterly by Scientific Research Publishing, Inc., USA.

Subscription rates:
Print: $79 per issue.
To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

SERVICES

Advertisements
Advertisement Sales Department, E-mail: service@scirp.org

Reprints (minimum quantity 100 copies)
Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.
E-mail: sub@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: ojgen@scirp.org

Scientific Research Publishing

Open Journal of Genetics, 2020, 10, 1-7
https://www.scirp.org/journal/ojgen
ISSN Online: 2162-4461
ISSN Print: 2162-4453

The True Idea of Mendel's Assumption regarding the Gene Is Rediscovered

Muying Zhou

The Central Hospital of Shandong Feicheng Coal-Mining Group Corporation, Taian, China
Email: fckzmy@sina.com

How to cite this paper: Zhou, M.Y. (2020) The True Idea of Mendel's Assumption regarding the Gene Is Rediscovered. *Open Journal of Genetics*, **10**, 1-7.
https://doi.org/10.4236/ojgen.2020.101001

Received: November 26, 2019
Accepted: December 20, 2019
Published: December 23, 2019

Abstract

The question of whether the "synthetic life" created by J.C. Venter was produced by DNA (the genome) also raises the question of whether the basic ideas behind modern genetics (germplasm = the hereditary material of the germ cells = genes) are true. In theory, however, whether genes can produce cells (including traits) should depend on Mendel's assumptions regarding genes and not on what people can argue or discuss at will. Consequently, the author studied Mendel's assumptions again. It turns out, unexpectedly, that the gene refers to a hereditary element controlling the specifications of the individual rather than the producer of the individual. That is to say, the gene is a facilitator, and the acceptor produces the individual by following the specifications set by the gene. This is the most significant genetics-related discovery since Mendel's death, Scientific facts, including Avery's experiment, consistently proved that this is true.

Keywords

Genes, Hereditary Material, Genetics, Template, Producer

1. Introduction

It is generally believed that "modern genetics", which regards Mendel as the "father of genetics", has a correct understanding of Mendel's assumptions regarding genes. However, a storm in 2010 made this idea shaken.

In 2010, J.C. Venter said that they had created a man-made genome (of *Mycoplasma mycoides*) and had used it to make "synthetic life" [1] [2]. This means: **genes** (genome) **are the producer of the individual** (cell). That is consistent with the "modern genetics" consensus: "germplasm: the hereditary material of the germ cells: genes" [3].

However, Dr. Gerald Joyce, an internationally renowned life scientist at the

Scripps Institute in California, said in the <*New York Times*>: "Dr. Venter copied the DNA from one species of bacteria and inserted it into another. The second bacteria made all the proteins and organelles in the so-called 'synthetic cell' by following the specifications implicit in the structure of the inserted DNA" [2]. This means: **genes** (genome) **are not the producer of the individual** (cell).

Now there are two contradictory judgments that cannot both be true.

Identifying the correct statement would directly determine the fate of "modern genetics". If genes are the producer of the individual (cell), then "modern genetics" is right; However, if genes are not the producer of the individual (cell), then genes cannot be the hereditary material, and the basis of "modern genetics" (in which genes are regarded as the hereditary material) is incorrect.

The "synthetic life" once praised as "the most important scientific research achievement in human history" was soon put aside. However, this occurred quietly, perhaps because the scientific community was afraid that people were paying too much attention or was afraid that people might ask why? The answer is that the academic community itself is in a state of confusion. No one directly stated that Dr. Venter's claim was ridiculous (In fact, it is absurd. If DNA could produce the cell, it should be a perpetual motion machine, and if DNA could produce the cell without consuming energy, then life on this planet would no longer need sunlight to reproduce). Additionally, no one has continued to espouse the claim of "synthetic life", and no one has questioned why this "highest achievement in molecular biology" has not won a Nobel Prize over the past 9 years. Neither Dr. Venter nor Dr. Joyce publicly continued to make their cases. Dr. Venter did not pose the following question to Dr. Joyce: If DNA is incapable of producing the cell, then is "modern genetics" incorrect? Nor did Dr. Joyce go further to claim that Dr. Venter's assertion was attributed to the basic, incorrect, idea behind "modern genetics".

This indicates that it may have been difficult for them to balance the beliefs of "modern genetics" with the facts of molecular biology. On the one hand, "modern genetics" has been a classical theory of great authority for 100 years; on the other hand, authoritative objective facts are important than any authority.

In theory, however, whether genes can produce cells (including traits) should depend on Mendel's assumptions regarding genes and not on what people can argue or discuss at will. Thus, we should go back to Mendel's original assumptions regarding genes to determine if genes, as defined by Mendel, are the producers of the individuals (including traits).

2. Mendel's Assumptions regarding the Gene

Mendel assumed: "if the tall variety contains in its germ cells something that makes the plants tall, and if the short variety carries something in its germ cells that makes the plants short". The "something" above is what was later called a "gene" [4].

From this assumption, Mendel did not think that the gene is the producer of tall (or short) or the producer of the plant. Instead, this assumption clearly tells us that the gene is the facilitator that makes the individual (plants) tall (or short). Remember that the gene is the facilitator rather than the producer. This is the basic idea of Mendel's assumption, and the key basis for judging which of the statements by Dr. Venter and Dr. Joyce is correct.

Apparently, Mendel considers that the gene does not produce tall (or short) or plant, but could frame plants to be tall (or short). Therefore, the gene is the element that frames and controls the specifications of the plant (individual).

Such facilitators that control product specifications are common in daily life. They often appear in the form of templates, drawings, molds and design schemes. For example, an aircraft factory contains drawings that can be used to make monoplanes and biplanes. One wing's drawing controls the specifications for one part (the wing) of the aircraft, and a set of drawings controls the specifications for the entire aircraft. In the same way, one gene (e.g., the tall-gene or short-gene) controls the specifications for one trait (tall or short, respectively), while an entire set of genes (genome) controls the specifications for an entire set of traits (e.g., a plant).

The facilitator is not the same as the producer, but it is a participant in the production of products having particular specifications. Being a facilitator logically implies the existence of its facilitating action's acceptor; without an acceptor, there would be no facilitator. In an aircraft factory, the acceptor is the aircraft production line. Thus, each aircraft is produced by the aircraft production line following the specifications defined in this aircraft's drawings. In Mendel's assumption, the acceptor can only exist in the gene-free part of the fertilized egg (today we know that this acceptor is the egg's transcriptase system [5]; however, this scientific discovery occurred in the second half of the 20th century. In Mendel's time there was no way to be sure), because the gene, as the facilitator, cannot also be the recipient of the facilitator. Each individual is produced by the gene-free part of the fertilized egg following the specifications defined in this individual's genes (genome). Therefore, just as an aircraft factory has two elements (the aircraft production line and aircraft design drawings), Mendel's assumption directly implies that the individual producer is made up of two elements. They are 1) the gene/genome (*i.e.*, facilitator) and 2) the producing operator (*i.e.*, recipient) located in gene-free part of the fertilized egg.

From the above analysis, we know that the original meaning of Mendel's assumptions regarding the gene was as follows: **the gene is a hereditary element controlling product (individual, trait) specifications. It is able to make the gene-free part of the fertilized egg to produce products (individual, trait) in accordance with the specifications set by it.** That is to say, the tall plants are produced by the gene-free part of the fertilized egg following the specifications defined in the genome containing the tall-gene, and the short plants are produced by the gene-free part of the fertilized egg following the specifications defined in the genome containing the short-gene.

Thus, we can see that the original meaning of Mendel's assumption is surprisingly consistent with Dr. Joyce's aforementioned statement.

Please see: the DNA synthesized by Dr. Venter was the genome of *Mycoplasma mycoides*. According to Mendel's thought it should be the hereditary element controlling individual specifications. Namely, it is able to make the gene-free part of the fertilized egg following the specifications it sets to produce a *M. mycoides* bacterium. This is exactly what Dr. Joyce states: "the second bacteria (an enucleated *Mycoplasma capricolum*, equivalent to the gene-free part of the fertilized egg) made all the proteins and organelles in the so-called 'synthetic cell' (belonging to *M. mycoides*) by following the specifications implicit in the structure of the inserted DNA".

This fully confirms that Mendel's assumptions regarding the gene were true, because Dr Joyce's words are supported by the scientific facts discovered over the past 76 years. Additionally, it also provides good circumstantial evidence that our interpretation (of Mendel's assumptions regarding the gene) is right.

3. Objective Facts Prove That Real-World Genes (DNA) Are Truly as Defined by Mendel, and There Is No Evidence of the Existence of the So-Called Genes (DNA) of "Modern Genetics" That Act as the Hereditary Materials or the Producers of Individuals

3.1. The First Evidence Comes from Avery's Experiment

In 1944, Avery *et al.* confirmed that genes are made of DNA and stated: "deoxyribonucleic acid (DNA) is capable of stimulating unencapsulated R variants of Pneumococcus Type II to produce a capsular polysaccharide" [6]. This told us that DNA (gene) is not the producer of a capsular polysaccharide (trait); instead, the gene only is the facilitator (stimulator) that makes unencapsulated R variants of Pneumococcus Type II (as a producing operator) to produce a capsular polysaccharide. Namely, the unencapsulated R variants of Pneumococcus Type II (as a producing operator) produced a capsular polysaccharide by following the specifications implicit in the structure of the DNA (discovered by Avery *et al*). Avery's claim is entirely consistent with Mendel's assumptions regarding the gene (and Dr Joyce's previous statement).

This experimental evidence indicates that real-world genes (DNA) are truly as defined by Mendel, and there is no evidence of the existence of the so-called genes (DNA) of "modern genetics" as the hereditary materials or the producers of individuals.

3.2. The Outline of Scientific (Molecular and Cellular Biology) Evidences

Following the Avery experiment, there was considerable progress in biochemistry (including the later appearance of molecular biology) and cytology (including the later appearance of cellular biology), especially in areas involving DNA.

The new findings included the transcription of DNA, the transcriptase and transcription factors, the base pairing principle, the genetic code, the discovery of RNAs and their respective functions, the confirmation of RNA's function producing protein, the confirmation of various proteins' (including various synthetic enzymes) functions, and the gradual and complete mastery of cell cycle knowledge. More than 100 Nobel Prize winners were among the scientists who made these contributions. These scientific achievements are now recorded in university textbooks and encyclopedias on the scientific fields (molecular biology, biochemistry, cytology, and cell biology).

In the vast scientific truth, it is enough to put forward the following four points: 1) **All the cells (including cell products) on the earth are produced by the DNA-free part (the transcriptase system) of the cell following the specifications implicit in the structure of cell's DNA (genome).** The production is accomplished using a natural, predetermined, automatic, cause-result continuous, procedural process (*i.e.*, the cell cycle) initiated by genome transcription, but all the activities can be attributed to the genome being transcribed; 2) All the cells of a particular individual are products of their first cell (fertilized egg); therefore, **each individual is produced by the DNA-free part of the fertilized egg (the transcriptase system) following the specifications implicit in the structure of the egg's DNA (genome).** Thus, the genome's transcription (by the egg's transcriptase system) is the root cause of the individual being produced; 3) The genome controls the specifications of all the products of the individual. Because DNA is the template for producing RNA, it first controls the specifications of the RNA, then RNA controls the specifications of the protein, and finally the protein controls the specifications of other organic substances, such as lipids and carbohydrates; and 4) DNA has no producing capacity. In the process of an individual's formation, DNA does not consume energy, does not do work, and does not build 3', 5'-phosphodiester or peptide bonds.

The first two points show that the scientific truth is exactly the same as Mendel's assumptions. Genes (DNA) are not the producers of the individuals (cells) but the facilitators. All the individuals (cells) are produced by the DNA-free part of the egg (the transcriptase system) following the specifications implicit in the structure of the DNA (genome) of the egg (the cell). Additionally, the last two points confirm that genes (DNA) are pure elements that frame and control the specifications of the product (individual, cell) in the form of a template.

As with Avery's experiment, these objective facts also indicate that real-world genes (DNA) are truly as defined by Mendel, and there is no evidence of the existence of the so-called genes (DNA) of "modern genetics" as the hereditary materials or the producer of the individual.

4. Conclusions

Now we can be sure that the true idea of Mendel's assumption regarding the gene has been rediscovered.

Rediscovering the original meaning of Mendel's assumption regarding the gene is the most significant discovery in genetics since Mendel's death because this meaning has never been recognized by the world. For about half a century before the birth of "modern genetics", few people knew of Mendel, and for the next 100 years, people were only exposed to the incorrect meaning provided by "Modern Genetics". This, of course, had consequences. The most serious consequence is that genes have become the basis on which "modern genetics" was established. A wrong theory would lead to absurdity. As a typical example, Dr. Venter believed he created "synthetic life" because he synthesized a genome.

Just as it was possible to research the sun, the earth, the moon, and the planets in the era of the Geocentric Theory, it was also possible to research genes, chromosomes, DNA, RNA, and transcriptase in the "Modern Genetics" era. However, the general direction of this science is wrong. If people insisted on following the Geocentric Theory, then because they think they have found the center of the solar system, they would never look for the center again, and they finally would never know where the center is. Similarly, if people insist on following "modern genetics", then because they think they have found the hereditary material (*i.e.* genes, while in reality the gene is only one element of the hereditary material), they would stop looking for the hereditary material, and they finally would never know what the hereditary material is. Consequently, people will never solve the mystery of life, and man-made life will never occur.

The correct theory of genetics is essential for, and a guarantee of, the normal development of genetics-based sciences.

Note: The author of this paper has published a paper on the same subject [7]. However, this paper focuses on "genes are templates rather than producers" and gives an in-depth analysis of the reasons why "modern genetics" treats genes as the hereditary material. Later, there were feedbacks that although they agreed with the conclusions of this article, their understanding of the consistency between Mendel's assumption and Dr Joyce's words was still vague. The author was prompted to write this article. When the basic framework of the paper is completed, the author found that the true idea of Mendel's assumption regarding the gene can be described in Dr. Joyce's language. With Dr. Joyce's reference, such a description is not only simple, but also particularly persuasive. It can be said that today, after the truth of the gene is revealed, the most appropriate expression language of the true idea of the Mendel's assumption regarding the gene could come on. Before, I understand the true idea of Mendel's assumption; however, how to persuade others always is very hard. Dr Joyce provided me with the most appropriate expression language. Only having been persuaded the world can realize that the true idea of Mendel's assumption regarding the gene has been rediscovered.

Acknowledgements

I thank Lesley Benyon, PhD, from Liwen Bianji, Edanz Group China

(www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Gibson, D.G., Grass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.Y., Algire, M.A., *et al.* (2010) Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science*, **329**, 52-56. https://doi.org/10.1126/science.1190719

[2] Wade, N. (2010) Researchers Say They Created a "Synthetic Cell". *New York Times*.

[3] Germplasm (2019) Merriam-Webster Dictionary of Encyclopaedia Britannica. https://www.merriam-webster.com/dictionary/germplasm

[4] Morgan, T.H. (1928) The Theory of the Gene. Yale University Press, New Haven, 2.

[5] Zhou, M.Y. (2018) The Answer to the "First Question" in Genetics: The Hereditary Material. *Open Access Library Journal*, **5**, e4645. https://doi.org/10.4236/oalib.1104645

[6] Avery, O.T., MacLeod, C.M. and McCarty, M. (1944) Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types. *Journal of Experiment Medicine*, **79**, 152. https://doi.org/10.1084/jem.79.2.137

[7] Zhou, M.Y. (2018) Mendel Actually Wants to Say: Genes Are Only One Element of the Hereditary Material. *Open Access Library Journal*, **5**, e4941. https://doi.org/10.4236/oalib.1104941

Scientific Research Publishing

Open Journal of Genetics, 2020, 10, 8-15
https://www.scirp.org/journal/ojgen
ISSN Online: 2162-4461
ISSN Print: 2162-4453

The Pattern of Occurrence of Cytosine in the Genetic Code Minimizes Deleterious Mutations and Favors Proper Function of the Translational Machinery

Bin Wang

Department of Chemistry, Marshall University, Huntington, WV, USA
Email: wangb@marshall.edu

How to cite this paper: Wang, B. (2020) The Pattern of Occurrence of Cytosine in the Genetic Code Minimizes Deleterious Mutations and Favors Proper Function of the Translational Machinery. *Open Journal of Genetics*, **10**, 8-15.
https://doi.org/10.4236/ojgen.2020.101002

Received: January 10, 2020
Accepted: February 29, 2020
Published: March 3, 2020

Abstract

The standard genetic code consists of 64 combinations of base triplets made from four different bases. The research aim of this study was to investigate the pattern of occurrence of cytosine in the genetic code. By exploring the base composition and sequence of all 64 codons, the author found some important features based on the instability of cytosine. Because cytosine undergoes spontaneous deamination that converts it into uracil, it is evolutionarily favorable to exclude cytosine from codons critical to the initiation and termination of translation. For amino acids that have one to three synonymous codons (also called synonyms), the frequency of occurrence of C in the first and second positions of their mRNA codons is significantly lower than the frequencies of A, U, and G. For mRNA codons that encode amino acids with four synonyms, the trend of base composition is opposite to those encoding amino acids with one to three synonyms; the instability of C could be inhibited or reduced via formation of hydrogen bonds with a G and/or with a protonated C, and the secondary structure of the resultant mRNA could be adjusted via the multiple synonymous alternates at the third position of their codons to facilitate the translation process. The overall pattern of occurrence for C in the genetic code not only minimizes deleterious mutations and favors proper function of the translational machinery by excluding C from certain positions within codons, but also allows the occurrence of genetic diversity via mutation by including C in less-critical positions.

Keywords

Genetic Code, Base Triplet, Synonyms, Cytosine Deamination,

Translation Mutations

1. Introduction

The standard genetic code is nearly universal, and consists of 64 combinations of base triplets made from four different bases—adenine (A), guanine (G), uracil (U), and cytosine (C). Since 61 of the 64 base triplets are used to encode only 20 amino acids, most amino acids are encoded by more than one codon. The remaining three triplets, called stop codons, designate the termination of translation [1]. To the author's knowledge, no study has investigated the pattern of occurrence of cytosine in the genetic code; it thus became the objective of this study. The author explored the base composition and sequence of all 64 codons, and inferred some important features in view of the instability of cytosine.

2. Methods

Since the genetic code is highly degenerate, meaning that most amino acids are encoded by more than one mRNA codon, the author divided the standard genetic codons into two groups: the base triplets encoding amino acids that have one to three synonymous codons (**Table 1**), and those amino acids with four synonymous codons (**Table 2**). Amino acids serine, leucine, and arginine each have six synonymous codons (also called synonyms); they are categorized as two-synonym plus four-synonym occurrences. The author determined the percentage (%) of A, U, G, and C at every position of the base triplet for mRNA codons with one to three synonyms (**Table 1**), and those with four synonyms (**Table 2**), respectively.

3. Results

The first feature is the absence of cytosine (C) in both the start (AUG, also the only codon for methionine) and stop codons (UAA, UAG, and UGA) of translation. The initiation and termination of translation are critical for protein synthesis; therefore, evolution has resulted in a higher frequency of the more stable A, U, and G to avoid a fatal malfunction in the translation process. Cytosine is also absent from the only codon for the amino acid tryptophan (UGG). The author infers that the absence of cytosine from the codons for methionine and tryptophan, neither of which has an alternate mRNA codon, is the result of evolutionary selection to avoid translation errors due to the spontaneous deamination of cytosine to uracil [2] [3] [4].

In contrast to the standard genetic code referred to above, mitochondrial genomes contain alternate start codons (e.g., AUA and AUU in humans, and GUG and UUG in prokaryotes). All vertebrate mitochondria use AGA and AGG as translation terminators. Mitochondrial mRNA from vertebrates and microorganisms use UGA to encode tryptophan rather than as a translation terminator,

Table 1. Analysis of the base triplets that encode the initiation and termination of translation, and those that encode amino acids with one to three synonymous codons. The genetic codons, and the amino acids encoded and their properties are from Berg *et al.* (2015) and Harris *et al.* (2016) [1] [6].

Amino Acid Encoded, Including the Property and Formula of Its Side Chain	mRNA Codon	% of Each Base for mRNA Codons with 1 - 3 Synonyms				
		1st Position (Left)	2nd Position (Middle)	3rd Position (Right)	1st and 2nd Positions	All Three Positions
Methionine (Met) Translation Start Codon hydrophobic $-CH_2CH_2SCH_3$	AUG					
Tryptophan (Trp) hydrophobic $-CH_2C_8H_6N$	UGG					
Lysine (Lys) positively charged $-CH_2CH_2CH_2CH_2NH_3^+$	AAA AAG	% of A 12/32 = 37.5%	% of A 16/32 = 50%	% of A 8/32 = 25%	% of A 28/64 = 43.8%	% of A 36/96 = 37.5%
Asparagine (Asn) polar $-CH_2CONH_2$	AAU AAC	% of U 12/32 = 37.5%	% of U 8/32 = 25%	% of U 8/32 = 25%	% of U 20/64 = 31.2%	% of U 28/96 = 29.2%
Arginine (Arg) positively charged $-CH_2CH_2CH_2NHC(NH_2)_2^+$	AGA AGG	% of G 4/32 = 12.5%	% of G 8/32 = 25%	% of G 8/32 = 25%	% of G 12/64 = 18.8%	% of G 20/96 = 20.8%
Serine (Ser) polar $-CH_2OH$	AGU AGC	% of C 4/32 = 12.5%	% of C 0/32 = 0%	% of C 8/32 = 25%	% of C 4/64 = 6.2%	% of C 12/96 = 12.5%
Tyrosine (Tyr) polar $-CH_2C_6H_4OH$	UAU UAC					
Leucine (Leu) hydrophobic $-CH_2CH(CH_3)_2$	UUA UUG					
Phenylalanine (Phe) hydrophobic $-CH_2C_6H_5$	UUU UUC					
Cysteine (Cys) polar $-CH_2SH$	UGU UGC					
Glutamic Acid (Glu) negatively charged $-CH_2CH_2COO^-$	GAA GAG					
Aspartic Acid (Asp) negatively charged $-CH_2COO^-$	GAU GAC					
Glutamine (Gln) polar $-CH_2CH_2CONH_2$	CAA CAG					
Histidine (His) polar/positively charged $-CH_2C_3H_3N_2$	CAU CAC					
Isoleucine (Ile) hydrophobic $-CH(CH_3)(CH_2CH_3)$	AUA AUU AUC					
Translation Stop Codon	UAA UAG UGA					

Table 2. Analysis of the base triplets that encode amino acids with four synonymous codons. The genetic codons, and the amino acids encoded and their properties are from Berg *et al.* (2015) and Harris *et al.* (2016) [1] [6].

Amino Acid Encoded, Including the Property and Formula of Its Side Chain	mRNA Codon	% of Each Base for mRNA Codons with Four Synonyms				
		1st Position (Left)	2nd Position (Middle)	3rd Position (Right)	1st and 2nd Positions	All Three Positions
Threonine (Thr) polar $-CHCH_3OH$	ACA					
	ACG					
	ACU					
	ACC					
Serine (Ser) polar $-CH_2OH$	UCA					
	UCG					
	UCU					
	UCC					
Valine (Val) hydrophobic $-CH(CH_3)_2$	GUA					
	GUG	% of A 4/32 = 12.5%	% of A 0/32 = 0%	% of A 8/32 = 25%	% of A 4/64 = 6.2%	% of A 12/96 = 12.5%
	GUU					
	GUC					
Glycine (Gly) hydrophobic $-H$	GGA	% of U 4/32 = 12.5%	% of U 8/32 = 25%	% of U 8/32 = 25%	% of U 12/64 = 18.8%	% of U 20/96 = 20.8%
	GGG					
	GGU					
	GGC	% of G 12/32 = 37.5%	% of G 8/32 = 25%	% of G 8/32 = 25%	% of G 20/64 = 31.2%	% of G 28/96 = 29.2%
Alanine (Ala) hydrophobic $-CH_3$	GCA					
	GCG					
	GCU	% of C 12/32 = 37.5%	% of C 16/32 = 50%	% of C 8/32 = 25%	% of C 28/64 = 43.8%	% of C 36/96 = 37.5%
	GCC					
Leucine (Leu) hydrophobic $-CH_2CH(CH_3)_2$	CUA					
	CUG					
	CUU					
	CUC					
Arginine (Arg) positively charged $-CH_2CH_2CH_2NHC(NH_2)_2^+$	CGA					
	CGG					
	CGU					
	CGC					
Proline (Pro) hydrophobic $-CH_2CH_2CH_2-$	CCA					
	CCG					
	CCU					
	CCC					

and vertebrate mitochondria use AUA for methionine rather than for isoleucine [1] [5] [6]. Again, C is absent from these critical codons. While the author will focus on the nucleic genetic code in the following discussion, it is noted that the pattern of occurrence for cytosine seems to be true for mitochondrial codons as well.

The right-hand column in **Table 1** ("All Three Positions" column) provides the total base composition, including total number and percentage of A, U, G, and C in the mRNA codons shown. Overall, A and U residues are more abundant than G and C residues in the codons for amino acids with one to three synonyms. Data presented in **Table 1** ("1st Position" column) provide the base composition at the 5'/left end of the base triplet of the mRNA codons studied. The frequencies of A and U are 37.5% each, whereas G and C residues are less frequent (12.5% each). At the second/middle base of the mRNA codons studied, the frequencies of A, U, and G are 50%, 25%, and 25%, respectively, as shown in

Table 1 ("2nd Position" column). Interestingly, C does not occur at the second position. At the 3'/right end of the base triplet of the mRNA codons studied (Table 1, "3rd Position" column), there is an equal abundance of A, U, G, and C (25% each).

For mRNA codons that encode amino acids with four synonyms, the trend of base composition is opposite to those encoding amino acids with one to three synonyms. As shown in Table 2 ("All Three Positions" column), C and G residues are more abundant than U and A residues for codons encoding amino acids with four synonyms. The frequencies of C and G at the first position of the mRNA codons studied (Table 2, "1st Position" column) are 37.5% each, whereas the frequencies of U and A are 12.5% each. At the second position of the mRNA codons studied (Table 2, "2nd Position" column), the frequencies of C, G, and U are 50%, 25%, and 25%, respectively; A does not occur at the second position. At the third position of the mRNA codons studied (Table 2, "3rd Position" column), there is an equal abundance of A, U, G, and C (25% each).

4. Discussion

Because cytosine is known to undergo spontaneous deamination into uracil, it is evolutionarily favorable to exclude cytosine from codons critical to the initiation or termination of translation. For amino acids that have one to three synonyms, the frequency of occurrence of C in the first and second positions (the root) of their mRNA codons is significantly lower than the frequencies of occurrence of A, U, and G (see Table 1, "1st and 2nd Positions" column). Furthermore, since the middle position of a base triplet is the most critical location for mRNA codon-tRNA anticodon interaction/binding [7] [8] [9] [10] [11], the complete absence of C from the second position that is observed for base triplets encoding amino acids with one to three synonyms is not surprising.

In Table 1, the only mRNA codons containing C in the root are those encoding histidine (CAU and CAC) and glutamine (CAA and CAG). If spontaneous deamination by hydrolysis occurs, histidine will be converted into tyrosine (UAU and UAC), and glutamine will be converted into a stop codon (UAA and UAG). Since histidine and tyrosine both have polar side chains, in theory, this C-to-U mutation may be less likely to introduce significant changes in a protein's structure or function. However, histidine is often found in active sites of enzymes because its imidazole ring-containing side chain is able to perform many different roles in catalysis, whereas tyrosine has a phenol-containing side chain [1] [6]. Therefore, the histidine-to-tyrosine mutation may allow for genetic variation. The C-to-U mutation within a glutamine codon would cause translation to stop. Because humans can synthesize enough glutamine, it is the most abundant nonessential amino acid in the human body; further studies are needed to determine the effects of the conversion of a glutamine codon into a stop codon on human health and on genetic diversity, although the loss of a protein is likely to have deleterious effects.

For amino acids that have four synonyms, the effects of an unstable C on

translation mutations may not be as deleterious as for amino acids with fewer synonyms, due to the high percentages of C and G in the root, and to the existence of multiple synonymous alternates at the third position of these codons. Frederico *et al.* demonstrated that the rate of hydrolytic deamination of cytosine in a double helix was approximately 140-fold slower than in single-stranded DNA at 37˚C [12]; this difference is mainly due to the decreased accessibility of the N3 and C4 positions in a cytosine that is paired to guanine via hydrogen bonds, blocking the attack from water. The mRNA codons encoding amino acids with four synonyms are CG-rich in the root (see **Table 2**, "1st and 2nd Positions" column), which indicates that they have the potential to inhibit or reduce cytosine deamination by folding upon themselves to form a C≡G double helix, and/or to form a hydrogen-bonded C^+-C i-motif if the RNA sequence is C-rich. (Note: Previous studies have proved the existence of i-motifs under physiological pH [13] [14].) Since CG-rich mRNA regions may form complicated secondary structures that hinder the translation process, producing the same amino acid no matter which of the four mRNA bases is in the third position allows the adjustment of the secondary structure of the resultant mRNA.

Table 2 shows that no A is present at the second position of base triplets encoding amino acids with four synonyms. Previous studies have indicated that the second base of mRNA codons determines the hydrophobicity of the encoded amino acids: The majority of codons for hydrophilic (polar and/or charged) amino acids have A in the second position; while the majority of codons for hydrophobic amino acids have U in the second position [7] [15] [16]. From **Table 1**, we can see that hydrophilic amino acids with one to three synonyms have A or G in the second position of their mRNA codons, while hydrophobic amino acids with one to three synonyms have U or G in their second position. From **Table 2**, we can see that hydrophilic amino acids with four synonyms have C or G in the second position of their mRNA codons, while hydrophobic amino acids have U or C or G in their second position. Since the majority of hydrophilic amino acids have two synonyms, it is reasonable that A is absent from the second position of mRNA codons that encode amino acids with four synonyms.

5. Conclusion

In summary, for amino acids that have one to three synonyms, the frequency of occurrence of C in the root of their mRNA codons is significantly lower than the frequencies of A, U, and G. For amino acids that have four synonyms, the instability of C may be inhibited or reduced via the formation of hydrogen bonds with a G and/or with a protonated C. In addition, the "new" secondary structure of the resultant mRNA could be adjusted via the multiple synonymous alternates in the codons' third positions, which could facilitate the translation process. The overall pattern of occurrence for C in the genetic code not only minimizes deleterious mutations and favors proper function of the translational machinery by excluding C from certain positions within codons, but also allows the occurrence of genetic diversity via mutation by including C in less-critical positions. Evolu-

tion is an excellent engineer.

Acknowledgements

This work is supported by the National Science Foundation under Award No. OIA-1458952. Any opinions, findings, and conclusions expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Conflicts of Interest

The author declares that she has no competing financial interests.

Data Availability Statement

All data generated and analyzed during this study are included in this published article.

References

[1] Berg, J.M., Tymoczko, J.L., Gatto Jr., G.J. and Stryer, L. (2015) Biochemistry. 8th Edition, W.H. Freeman & Company, New York, NY.

[2] Nabel, C.S., Manning, S.A. and Kohli, R.M. (2012) The Curious Chemical Biology of Cytosine: Deamination, Methylation, and Oxidation as Modulators of Genomic Potential. *ACS Chemical Biology*, **7**, 20-30. https://doi.org/10.1021/cb2002895

[3] Poole, A., Penny, D. and Sjoberg, B.M. (2001) Confounded Cytosine! Tinkering and the Evolution of DNA. *Nature Reviews Molecular Cell Biology*, **2**, 147-151. https://doi.org/10.1038/35052091

[4] Levy, M. and Miller, S.L. (1998) The Stability of the RNA Bases: Implications for the Origin of Life. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 7933-7938. https://doi.org/10.1073/pnas.95.14.7933

[5] Swire, J., Judson, O.P. and Burt, A. (2005) Mitochondrial Genetic Codes Evolve to Match Amino Acid Requirements of Proteins. *Journal of Molecular Evolution*, **60**, 128-139. https://doi.org/10.1007/s00239-004-0077-9

[6] Harris, D.C. and Lucy, C.A. (2016) Quantitative Chemical Analysis. 9th Edition, W.H. Freeman & Company, New York, NY.

[7] Lehmann, J. and Libchaber, A. (2008) Degeneracy of the Genetic Code and Stability of the Base Pair at the Second Position of the Anticodon. *RNA*, **14**, 1264-1269. https://doi.org/10.1261/rna.1029808

[8] Auffinger, P. and Westhof, E. (2001) An Extended Structural Signature for the tRNA Anticodon Loop. *RNA*, **7**, 334-341. https://doi.org/10.1017/S1355838201002382

[9] Rumer, Y.B. (2016) Translation of "Systematization of Codons in the Genetic Code [I]" by Yu. B. Rumer (1966). *Philosophical Transactions of The Royal Society A*, **374**, Article ID: 20150446. https://doi.org/10.1098/rsta.2015.0446

[10] Rumer, Y.B. (2016) Translation of "Systematization of Codons in the Genetic Code [II]" by Yu. B. Rumer (1968). *Philosophical Transactions of The Royal Society A*, **374**, Article ID: 20150447. https://doi.org/10.1098/rsta.2015.0447

[11] Rumer, Y.B. (2016) Translation of "Systematization of Codons in the Genetic Code [III]" by Yu. B. Rumer (1969). *Philosophical Transactions of The Royal Society A*,

374, Article ID: 20150448. https://doi.org/10.1098/rsta.2015.0448

[12] Frederico, L.A., Kunkel, T.A. and Shaw, B.R. (1990) A Sensitive Genetic Assay for the Detection of Cytosine Deamination: Determination of Rate Constants and the Activation Energy. *Biochemistry*, **29**, 2532-2537. https://doi.org/10.1021/bi00462a015

[13] Wright, E.P., Huppert, J.L. and Waller, Z.A.E. (2017) Identification of Multiple Genomic DNA Sequences Which Form I-motif Structures at Neutral pH. *Nucleic Acids Research*, **45**, 2951-2959. https://doi.org/10.1093/nar/gkx090

[14] Zeraati, M., Langley, D.B., Schofield, P., Moye, A.L., Rouet, R., Hughes, W.E., Bryan, T.M., Dinger, M.E. and Christ, D. (2018) I-motif DNA Structures Are Formed in the Nuclei of Human Cells. *Nature Chemistry*, **10**, 631-637. https://doi.org/10.1038/s41557-018-0046-3

[15] Chiusano, M.L., Alvarez-Valin, F., Di Giulio, M., D'Onofrio, G., Ammirato, G., Colonna, G. and Bernardi, G. (2000) Second Codon Positions of Genes and the Secondary Structures of Proteins. Relationships and Implications for the Origin of the Genetic Code. *Gene*, **261**, 63-69. https://doi.org/10.1016/S0378-1119(00)00521-7

[16] Copley, S.D., Smith, E. and Morowitz, H.J. (2005) A Mechanism for the Association of Amino Acids with Their Codons and the Origin of the Genetic Code. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 4442-4447. https://doi.org/10.1073/pnas.0501049102

Call for Papers

Open Journal of Genetics

ISSN: 2162-4453 (Print), 2162-4461 (Online)
https://www.scirp.org/journal/ojgen

Editor-in-Chief

Editorial Board

Open Journal of Genetics (OJGen) is an international journal dedicated to the latest advancements of Genetics. The goal of this journal is to provide a platform for scientists and academicians all over the world to promote, share, and discuss various new issues and developments in different areas of Genetics. All manuscripts must be prepared in English, and are subject to a rigorous and fair peer-review process. Accepted papers will immediately appear online followed by printed hard copy. The journal publishes original papers including but not limited to the following fields:

- Classical and Developmental Genetics
- Conservation and Ecological Genetics (including Behavioural Genetics)
- Evolutionary and Population Genetics (except Human)
- Genetic Engineering
- Genetics of Plants and Animals in Agriculture (including Quantitative Genetics)
- Genomics
- Human and Medical Genetics (including Quantitative Genetics, Population Genetics, Psychiatric Genetics)
- Immunogenetics
- Microbial Genetics
- Molecular Genetics (including Gene Characterization and Polymorphisms, Control of Gene Expression and Epigenetics)
- Technical Tips and Improvement

We are also interested in: 1) Short Reports—2-5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data; 2) Book Reviews—Comments and critiques.

Website and E-Mail

https://www.scirp.org/journal/ojgen E-mail: ojgen@scirp.org

What is SCIRP?

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

What is Open Access?

All original research papers published by SCIRP are made freely and permanently accessible online immediately upon publication. To be able to provide open access journals, SCIRP defrays operation costs from authors and subscription charges only for its printed version. Open access publishing allows an immediate, worldwide, barrier-free, open access to the full text of research papers, which is in the best interests of the scientific community.

- High visibility for maximum global exposure with open access publishing model
- Rigorous peer review of research papers
- Prompt faster publication with less cost
- Guaranteed targeted, multidisciplinary audience

Scientific Research Publishing

Website: https://www.scirp.org
Subscription: sub@scirp.org
Advertisement: service@scirp.org