

Full Research Paper

A Statistical Analysis of the Robustness of Alternate Genetic Coding Tables

Mehmet Levent Kurnaz¹, **Tugce Bilgin**² and **Isil Aksan Kurnaz**^{2,*}

1 Physics Department, Bogazici University Bebek, 34342 Istanbul, Turkey. E-mail: levent.kurnaz@boun.edu.tr

2 Department of Genetics and Bioengineering, Yeditepe University Kayisdagi, 34755 Istanbul, Turkey. E-mail: tugce_bilgin@yahoo.co.uk. E-mail: iakurnaz@yeditepe.edu.tr

* Author to whom correspondence should be addressed.

Received: 18 December 2007; in revised form: 25 February 2008 / Accepted: 11 April 2008 / Published: 2 May 2008

Abstract: The rules that specify how the information contained in DNA is translated into amino acid “language” during protein synthesis are called “the genetic code”, commonly called the “Standard” or “Universal” Genetic Code Table. As a matter of fact, this coding table is not at all “universal”: in addition to different genetic code tables used by different organisms, even within the same organism the nuclear and mitochondrial genes may be subject to two different coding tables. Results In an attempt to understand the advantages and disadvantages these coding tables may bring to an organism, we have decided to analyze various coding tables on genes subject to mutations, and have estimated how these genes “survive” over generations. We have used this as indicative of the “evolutionary” success of that particular coding table. We find that the “standard” genetic code is not actually the most robust of all coding tables, and interestingly, Flatworm Mitochondrial Code (FMC) appears to be the highest ranking coding table given our assumptions. Conclusions It is commonly hypothesized that the more robust a genetic code, the better suited it is for maintenance of the genome. Our study shows that, given the assumptions in our model, Standard Genetic Code is quite poor when compared to other alternate code tables in terms of robustness. This brings about the question of why Standard Code has been so widely accepted by a wider variety of organisms instead of FMC, which needs to be addressed for a thorough understanding of genetic code evolution.

Keywords: genetic code, evolution, robustness, statistical analysis.

1. Introduction

How the genetic code evolved has been a matter of interest for many researchers over the past decades – Crick [1] had postulated the coevolution and frozen accident hypotheses, where similar amino acids would end up using similar codons as a result of coevolution of coding tables and genes, and remain “frozen” at an optimum coding that reduces deleterious effects of mutations (reviewed in [2]). One of the important properties of a genetic code is its *robustness* to error, which means that if a mutation occurs in a gene, the amino acid substitution ideally renders a functionally similar protein, thus a robust code reduces the deleterious effects of mutations. Thus one would at first sight assume that the coding table that has been adopted by a wider range of organisms would appear more robust, which has been the basic premise behind our analysis.

The genetic information about the individuals is stored in the DNA, which make up the genes. DNA is made up of different monomers, or nucleotides, containing one of the four heterocyclic bases: adenine (A), guanine (G), cytosine (C) and thymine (T). Genes use triplet codes (“codons”) to translate the information into proteins – each of the 20 amino acids is coded by three-base combinations (Figure 1). The Genetic Code Tables summarize how this codon assignment is made, saving three codons to signal “STOP” for protein synthesis machinery, i.e. 20 amino acids are encoded by 61 different codons. There are various exceptions to this Universal/Standard Coding Table, however – for instance vertebrate and invertebrate mitochondria use different coding tables for their own genes, as do Ciliates (Table 1). The alternate coding tables are believed to have arisen from the evolution of the standard genetic code through codon reassignments, and most studies on possible mechanisms of this evolution start out by the assumption that the changes resulting in codon reassignment would be strongly disadvantageous and consequently get eliminated from the system [3,4]. Using a similar assumption, our present study aims to compare the possible “evolutionary” advantages of these different genetic codes in terms of robustness and resilience to mutations.

In our study, genes or “individuals” are represented by bit-strings which are 32 bits long and are initially set to zero. Each bit represents a given age or generation: as the individual reproduces we move down on the bit-string. Bits which are set to zero represent that no deleterious mutations happened at that age. However, if a bit is set to one, it means that the individual suffers a severe mutation at that generation and its probability of survival or viability is compromised. This is based on previous reports that a lineage of organisms where mutations result in chemically conserved amino acid substitutions may actually have higher survivability as compared to those with a less conservative code [5,6], and according to the error-minimization hypotheses, the universal or standard genetic code has evolved an inverse relationship between the severity and frequency of these alterations [5,7]. We have previously used this model to show the optimal number of amino acids that could be encoded by 64 codons without affecting survivability of populations due to deleterious mutations [8]. In this initial set of analyses, we make two simplified assumptions: first of all, we assumed that nucleotide substitutions occur at similar frequencies (current work is integrating unequal substitution rates; unpublished data). Secondly, we assume that any change in amino acid composition would be deleterious, hence we do not incorporate similarity matrices for the purposes of simplification in this present study (ongoing work is incorporating BLOSUM matrices, without significant alterations in our findings; unpublished data).

Here, we have analyzed a variety of genes from different organisms against 12 different coding tables. Our analysis is based on the fundamental assumption that if the gene being analyzed is coding for an

essential component of the cells in that organism, such as integrity, metabolism, or replication of DNA, it becomes very important that the gene remain functional in order for the organism to survive. The underlying assumption is that the mutations which render this particular gene completely inactive would mean that the individual would not “survive” [8]. Thus, over a number of generations, we could analyze what the survivability outcome is with respect to the entire population – since the mutation is considered in the light of the particular coding table analyzed, the better the survivability, the more robust the coding table (for details, see Methods).

Figure 1. The “Universal Genetic Code Table”, adopted from Introduction to Biology, Campbell and Reece (6th Ed, 2002). START codon (AUG) encodes for Methionine (Met, M), and the three STOP codons are indicated (UAA, UAG, UGA).

1st Base	2nd Base	3rd Base	amino acid	(single letter)		
T	T	T	Phe	F		
		C	Phe	F		
		A	Leu	L		
		G	Leu	L		
	C	T	T	Ser	S	
			C	Ser	S	
			A	Ser	S	
		G	T	Ser	S	
			C	Tyr	Y	
			A	Tyr	Y	
	A	A	A	STOP	*	
			G	STOP	*	
T			Cys	C		
G		C	Cys	C		
		A	STOP	*		
		G	Trp	W		
C	T	T	Leu	L		
		C	Leu	L		
		A	Leu	L		
		G	Leu	L		
	C	T	Pro	P		
		C	Pro	P		
		A	Pro	P		
		G	Pro	P		
		A	T	His	H	
			C	His	H	
	A		Gln	Q		
	G		Gln	Q		
	G		T	Arg	R	
			C	Arg	R	
		A	Arg	R		
		G	Arg	R		
		A	T	T	Ile	I
				C	Ile	I
A	Ile			I		
G	Met			M		
C	T		Thr	T		
	C		Thr	T		
	A		Thr	T		
	G		Thr	T		
	A		T	Asn	N	
			C	Asn	N	
A			Lys	K		
G			Lys	K		
G		T	Ser	S		
		C	Ser	S		
	A	Arg	R			
	G	Arg	R			
	G	T	T	Val	V	
			C	Val	V	
A			Val	V		
G			Val	V		
C		T	Ala	A		
		C	Ala	A		
		A	Ala	A		
		G	Ala	A		
		A	T	Asp	D	
			C	Asp	D	
A			Glu	E		
G			Glu	E		
G			T	Gly	G	
			C	Gly	G	
		A	Gly	G		
		G	Gly	G		

Table 1. Comparison of various genetic coding tables (information accessed from NCBI Entrez).

Abbrev. Standard	Code Table Standard Code	Differences from Standard
VMC	The Vertebrate Mitochondrial Code	Code 2 Std AGA stop * Arg R AGG stop * Arg R AUA Met M Ile I UGA Trp W stop *
YMC	The Yeast Mitochondrial Code	Code 3 Std AUA Met M Ile I CUU Thr T Leu L CUC Thr T Leu L CUA Thr T Leu L CUG Thr T Leu L UGA Trp W stop *
MSC	The Mold, Protozoan, and Coelenterate Mitochondrial Code and the Mycoplasma/Spiroplasma Code	CGA absent Arg R CGC absent Arg R Code 4 Std UGA Trp W stop *
IMC	The Invertebrate Mitochondrial Code	Code 5 Std AGA Ser S Arg R AGG Ser S Arg R AUA Met M Ile I UGA Trp W stop *
CDH	The Ciliate, Dasycladacean and Hexamita Nuclear Code	Code 6 Std UAA Gln Q stop * UAG Gln Q stop *
EMC	The Echinoderm Mitochondrial Code	Code 9 Std AAA Asn N Lys K AGA Ser S Arg R AGG Ser S Arg R UGA Trp W stop *
ENC	The Euplotid Nuclear Code	Code 10 Std UGA Cys C stop *
AYNC	The Alternative Yeast Nuclear Code	Code 12 Std CUG Ser S Leu L

Table 1. cont.

		Code 13 Std
		AGA Gly G Arg R
AMC	The Ascidian Mitochondrial Code	AGG Gly G Arg R
		AUA Met M Ile I
		UGA Trp W stop *
		Code 14 Std
		AAA Asn N Lys K
FMC	The Flatworm Mitochondrial Code	AGA Ser S Arg R
		AGG Ser S Arg R
		UAA Tyr Y stop *
		UGA Trp W stop *
BNC	Blepharisma Nuclear Code	Code 10 Std
		UAG Gln Q stop *

Our results show that the “Universal Genetic Code” is actually sub-optimal in terms of robustness in this simplified analysis, and FMC appears to function significantly better in protecting the genes against mutations described in our study. For ciliate and hexamite representative genes only, The Ciliate, Dasycladacean and Hexamita Nuclear Code (CDH) appears to be on a par with FMC in terms of robustness, while Yeast (YMC) and Vertebrate Mitochondrial (VMC) codes are unsuccessful. It is rather puzzling that a relatively poor-performing Standard Code Table has been adopted by such a wide variety of organisms, and further analyses need to be performed in order to thoroughly understand the nature of the genetic codes. It should be noted, however, that differences in nucleotide substitution rates, and various amino acid substitution matrices should be incorporated in a larger study (work still ongoing), however our preliminary results indicate that the overall profile of robustness among coding tables do not significantly change (unpublished data). It should also be noted that the initial environment when the coding tables were still diverging was significantly different than the conditions today, and responses of populations to mutations could be similarly different, and some mutations could perhaps have been allowed. This study should therefore be further improved in order to consider many aspects, but should be seen as an initial step towards such an improvement.

2. Results and Discussion

A previous study [8] had studied how an *in silico* population survived over generations, by calculating the probability of “survival” upon random mutations of an essential gene – the so-called “human cytokine” gene – where a mutation that renders the protein non-functional resulted in death of the organism. The results in that study were rather interesting, showing that the optimum number of amino acids that could be encoded by the coding table that resulted in optimum survival of the population was indeed 22, rather than the 20 amino acids normally found in the Universal Coding Table (Figure 1, [8]).

This result by itself was rather intriguing, taken together with the fact that some genetic code reassignments and expansions of the coding tables are still ongoing [9]. This has led us to the question of the performance of the alternate coding tables. Using the same statistical analysis, we wished to address

whether the universal coding table was even slightly more robust than the alternate tables in terms of resilience against mutations.

To that end, we have analyzed several genes that are either ubiquitous or important for the integrity and functionality of cells of the body in humans and primates; such as actins, which are highly conserved proteins involved in cell motility and maintenance of the cytoskeleton, and tubulin isoforms, which are the main components of the microtubular network and functionally important for cellular integrity as well as mitosis (see Table 2 for a comprehensive list of genes and NCBI accession numbers).

Table 2. Representative genes that are encoded by different coding tables were used in this study.

Coding Table	GenBank Accession Number	Gene Name / Explanation
YMC	X69431	<i>Kluyveromyces thermotolerans</i> cox 2
YMC	X69430	<i>Candida glabrata</i> cox 2
YMC	X02439	<i>Hansenula saturnus</i> cox 2
YMC	AF442220	<i>Kluyveromyces lodderae</i> cox 2 (truncated)
YMC	KLU75348	<i>Kluyveromyces lactis</i> ATPase 9
YMC		Mitochondrion <i>Candida glabrata</i> average *
YMC		Mitochondrion <i>Kluyveromyces thermotolerans</i> average *
YMC		Mitochondrion <i>Kluyveromyces lactis</i> average *
IMC	AF329059; CDS 34-618	<i>Haementeria tuberculifera</i> NADH dehydrogenase subunit I (ND1) gene, partial cds; mitochondrial.
IMC	DQ202128; CDS 32-520	<i>Drosophila stalker</i> voucher NADH dehydrogenase subunit 2 (NADH2) gene, partial cds; mitochondrial.
IMC	AB275882	<i>Caenorhabditis</i> mitochondrial ND5 gene for NADH dehydrogenase
IMC	X99667	<i>Drosophila melanogaster</i> mRNA for mitochondrial ATPase synthase, subunit d
IMC	DROMTM2A	<i>Drosophila melanogaster</i> NADH dehydrogenase 3
IMC	AF164587	<i>Drosophila melanogaster</i> NADH dehydrogenase subunit 1
IMC	S76764	<i>Drosophila melanogaster</i> ND5, NADH dehydrogenase subunit 5
IMC		<i>Caenorhabditis elegans</i> average *
IMC		<i>Drosophila melanogaster</i> average *
VMC	NM_002488	<i>Homo sapiens</i> NADH hydrogenase 1 alpha subcomplex 2
VMC	BC128726	<i>Rattus norvegicus</i> ATP synthase, H ⁺ -transporting, mitochondrial F0 complex, subunit c
VMC	BC010318	<i>Mus musculus</i> PEP carboxykinase 2, mitochondrial
VMC	X79547	<i>Equus caballus</i> mitochondrial DNA complete sequence NADH dehydrogenase
VMC	NM_001079924	<i>Pan troglodytes</i> NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, NDUFA4, mitochondrial

Table 2. cont.

VMC	PTU12706	<i>Pan troglodytes</i> Ptr5 mitochondrion cytochrome oxidase subunit II (COII) gene
VMC	NM_008617	<i>Mus musculus</i> malate dehydrogenase 2, NAD (mitochondrial) (Mdh2)
VMC	NM_029696	<i>Mus musculus</i> malate dehydrogenase 1B, NAD (soluble) (Mdh1b), mRNA, mitochondrial
VMC	NM_008618	<i>Mus musculus</i> malate dehydrogenase 1, NAD (soluble) (Mdh1)
VMC	NM_010344	<i>Mus musculus</i> glutathione reductase 1 (Gsr)
VMC	NM_001009329	<i>Felis catus</i> cytosolic malate dehydrogenase (MDH)
VMC		<i>Equus caballus</i> mitochondrion average *
VMC		<i>Pan troglodytes</i> mitochondrion average *
VMC		<i>Mus musculus</i> mitochondrion average *
FMC	AJ621238	<i>Echinococcus granulosus</i> malate dehydrogenase
FMC	AF188122	<i>Clonorchis sinensis</i> cytochrome oxidase subunit 1
FMC	DQ402037	<i>Echinococcus granulosus</i> NADH dehydrogenase subunit 1 (ND1) gene, partial cds; mitochondrial.
FMC	AY147416	<i>Echinococcus granulosus</i> thioredoxin glutathione reductase
FMC		Flatworm (<i>E. granulosus</i>) mitochondria average*
ENC	AY124990	<i>Euplotes aediculatus</i> alpha-2 platein precursor, gene, complete cds
ENC	X71353	<i>Euplotes octocarinatus</i> gamma tubulin
ENC	EF030059	<i>Euplotes nobilii</i> pheromone En-6
ENC	DQ866998	<i>Euplotes nobilii</i> heat shock protein 70
ENC	Y09551	<i>Euplotidae crassus</i> gamma tubulin 2
ENC	AF273753	<i>Euplates vannus</i> actin1
ENC	AY295877	<i>Euplates focardii</i> HSP70
ENC	DQ864704	<i>Euplotes octocarinatus</i> beta2 tubulin
ENC	S72098	<i>Euplates focardii</i> beta tubulin
ENC	J04533	<i>Euplotidae crassus</i> actin
ENC		<i>Euplotes focardii</i> average *
ENC		<i>Euplotes vannus</i> average *
CDH	AY293806	<i>Paraurostyla weissei</i> macronuclear DNA polymerase alpha gene, complete cds
CDH	HIU37081	<i>Hexamita inflata</i> elongation factor 1 alpha gene, partial cds.
CDH	Z11836	<i>Stylonychia lemnae</i> gene for DNA Polymerase II.
CDH	AY008386	<i>Urostyla grandis</i> macronuclear type II DNA polymerase alpha gene, complete cds.
CDH	X57926	<i>Stylonychia lemnae</i> EF1
CDH	AF194336	<i>Stylonychia lemnae</i> micronuclear DNA polymerase
CDH	XM_001032213	<i>Tetrahymena thermophila</i> EF1
CDH	XM_001031057	<i>Tetrahymena thermophila</i> EFG
CDH		<i>Tetrahymena thermophila</i> average *

Table 2. cont.

CDH		<i>Stylonychia lemnae</i> average *
MSC	X65223	<i>Trichophyton rubrum</i> NADH 4L
MSC	X65223	<i>Trichophyton rubrum</i> cox 2
MSC	X65223	<i>Trichophyton rubrum</i> cox 1
MSC	NEUMTCOIJ	<i>Neurospora crassa</i> cox 2
MSC	AY548157	<i>Neurospora crassa</i> NADH dehydrogenase 1
MSC		<i>Neurospora crassa</i> average *
MSC		<i>Trichophyton rubrum</i> average *
Std	NM_001614	Human actin, gamma1
Std	AB062393	Human tubulin-beta
Std	AF141347	Human tubulin-alpha
Std	HUMACTA1	Human actin-beta
Std	AB292109	<i>Equus caballus</i> HSP70A8
Std	AB292108	<i>Equus caballus</i> EF1A1
Std	NM_001081838	<i>Equus caballus</i> actin beta
Std	X69884	<i>Equus caballus</i> CD2
Std	NM_001009165	<i>Pan troglodytes</i> EF1 alpha1
Std	NM_001009945	<i>Pan troglodytes</i> actin beta
Std	NM_001034095	<i>Pan troglodytes</i> tubulin alpha 1b
Std	NM_001045509	<i>Pan troglodytes</i> tubulin
Std	NM_001098544	<i>Pan troglodytes</i> tubulin alpha 1a
Std	NM_001098572	<i>Pan troglodytes</i> alpha 1
Std	AF091101	<i>Mus musculus</i> dUTPase
Std	MUSHSC70T	<i>Mus musculus</i> Hsc70T
Std	NM_007906	<i>Mus musculus</i> EF1 alpha 2
Std	NM_007393	<i>Mus musculus</i> actin beta
Std	NM_009609	<i>Mus musculus</i> actin gamma1
Std	NM_134024	<i>Mus musculus</i> tubulin gamma 1
Std	NM_009984	<i>Mus musculus</i> cathepsin L
Std	NM_013486	<i>Mus musculus</i> CD2
Std	NM_001009326	<i>Felis catus</i> EF1 alpha
Std	NM_001009841	<i>Felis catus</i> CD2
Std	EF407948	<i>Fasciola hepatica</i> cathepsin L mRNA (flatworm)
Std	EF201934	<i>Taenia asiatica</i> calcineurin B (flatworm)
Std	DQ256465	<i>Schistosoma mansoni</i> cathepsin-like protein CD2 (flatworm)
Std	EF199625	<i>Taenia solium</i> dUTPase (flatworm)
Std		Human average *
Std		<i>Equus caballus</i> average *
Std		<i>Pan troglodytes</i> average *
Std		<i>Mus musculus</i> average *
Std		<i>Felis catus</i> average *

* (based on genome-based codon usage frequencies obtained from <http://www.kazusa.or.jp/codon/>)

Figure 2. Comparison of various coding tables with respect to robustness for selected human and primate genes that are normally subject to Standard Genetic Code. “Average” genes represent hypothetical and idealized genes constructed using average codon usage frequencies (see Methods); accession numbers of the genes are listed in Table 2. Coding table abbreviations are given in Table 1.

	Human cytokine		Human actin γ 1		Human tubulin β		Human tubulin α		Human actin β		Human Average
FMC	1.4053	FMC	1.431	FMC	1.4462	FMC	1.4325	FMC	1.4694	FMC	1.4083
EMC	1.4163	CDH	1.4434	EMC	1.4644	EMC	1.4533	BHC	1.4831	EMC	1.4207
IMC	1.4319	BHC	1.4447	IMC	1.4786	CDH	1.4586	CDH	1.4831	IMC	1.4439
EHC	1.4409	EMC	1.4509	CDH	1.4797	BHC	1.4618	EMC	1.4831	CDH	1.4501
BHC	1.4685	EHC	1.4522	BHC	1.4808	EHC	1.4639	EHC	1.4831	BHC	1.4558
MSC	1.4755	MSC	1.4547	EHC	1.4951	MSC	1.4725	Standard	1.497	EHC	1.4596
AMC	1.4779	Standard	1.4585	AMC	1.4996	Standard	1.4757	IMC	1.497	MSC	1.4658
CDH	1.4779	IMC	1.4611	MSC	1.4996	AMC	1.4789	MSC	1.497	AMC	1.4697
Standard	1.4826	AMC	1.4713	Standard	1.5030	IMC	1.4811	AMC	1.5182	Standard	1.4716
VMC	1.5017	VMC	1.4817	VMC	1.5097	VMC	1.4985	VMC	1.5254	VMC	1.4971
YMC	1.5638	YMC	1.5461	YMC	1.5679	YMC	1.5586	YMC	1.562	YMC	1.5543
AYHC	1.5793	AYHC	1.5886	AYHC	1.5787	AYHC	1.5644	AYHC	1.6157	AYHC	1.5804

	<i>P. troglodytes</i> EF1 α 1		<i>P. troglodytes</i> actin β		<i>P. troglodytes</i> tubulin α 1b		<i>P. troglodytes</i> tubulin		<i>P. troglodytes</i> tubulin α 1a		<i>P. troglodytes</i> HSP90 α 1		<i>P. troglodytes</i> average
FMC	1.424201	FMC	1.438459	FMC	1.43258	FMC	1.448355	FMC	1.436708	FMC	1.53521	FMC	1.3957
EMC	1.436237	CDH	1.450991	EMC	1.453393	EMC	1.466586	EMC	1.457607	EMC	1.55365	EMC	1.4097
CDH	1.461788	BHC	1.452253	CDH	1.459722	IMC	1.480756	CDH	1.458664	CDH	1.579758	IMC	1.4294
BHC	1.462824	EMC	1.456048	BHC	1.460781	CDH	1.481854	BHC	1.46184	BHC	1.588614	CDH	1.4341
IMC	1.46386	EHC	1.459857	EHC	1.465027	BHC	1.482954	EHC	1.463964	EHC	1.593817	BHC	1.4392
EHC	1.463861	IMC	1.462405	MSC	1.473574	EHC	1.497358	MSC	1.472502	MSC	1.596055	EHC	1.4463
MSC	1.465937	MSC	1.462405	Standard	1.476798	AMC	1.501833	Standard	1.475722	Standard	1.598298	MSC	1.4496
Standard	1.471148	Standard	1.46752	IMC	1.478953	MSC	1.501833	AMC	1.481113	IMC	1.599796	AMC	1.4529
AMC	1.476387	AMC	1.47783	AMC	1.482195	Standard	1.505201	IMC	1.481113	AMC	1.613381	Standard	1.4572
VMC	1.491202	VMC	1.485632	VMC	1.498562	VMC	1.511973	VMC	1.498561	VMC	1.636441	VMC	1.4818
AYHC	1.538128	YMC	1.554532	AYHC	1.56332	AYHC	1.577494	AYHC	1.564495	AYHC	1.676098	YMC	1.5491
YMC	1.548235	AYHC	1.581465	YMC	1.572754	YMC	1.583558	YMC	1.568025	YMC	1.689947	AYHC	1.5697

Figure 3. Comparison of various coding tables with respect to robustness for selected genes from cat, horse and mouse that are normally subject to Standard Genetic Code. “Average” genes represent hypothetical and idealized genes constructed using average codon usage frequencies (see Methods); accession numbers of the genes are listed in Table 2. Coding table abbreviations are given in Table 1.

	<i>E. caballus</i> HSP70A8		<i>E. caballus</i> EF1A1		<i>E. caballus</i> actin β		<i>E. caballus</i> CD2		<i>E. caballus</i> average
FMC	1.449929	FMC	1.421215	FMC	1.443453	FMC	1.396565	FMC	1.4170
EMC	1.461711	EMC	1.433215	CDH	1.454781	EMC	1.406941	EMC	1.4308
CDH	1.462452	CDH	1.45972	BHC	1.456048	CDH	1.474352	IMC	1.4539
BHC	1.467655	IMC	1.460753	EMC	1.46113	IMC	1.475747	CDH	1.4577
EHC	1.479649	BHC	1.460754	EHC	1.463681	BHC	1.485572	BHC	1.4639
MSC	1.481159	EHC	1.461788	IMC	1.464959	EHC	1.49265	EHC	1.4677
Standard	1.481914	MSC	1.463861	MSC	1.466239	MSC	1.498348	MSC	1.4731
AMC	1.492552	Standard	1.469061	Standard	1.471373	Standard	1.499777	AMC	1.4760
IMC	1.492552	AMC	1.471148	AMC	1.479126	AMC	1.520009	Standard	1.4799
VMC	1.514175	VMC	1.488009	VMC	1.488246	VMC	1.54514	VMC	1.5026
AYHC	1.541878	YMC	1.540365	YMC	1.558736	AYHC	1.573995	YMC	1.5654
YMC	1.576167	AYHC	1.541485	AYHC	1.585784	YMC	1.603708	AYHC	1.6006

	<i>M. musculus</i> dUTPse		<i>M. musculus</i> Hsc70T		<i>M. musculus</i> EF1 α2		<i>M. musculus</i> actin β		<i>M. musculus</i> actin γ1		<i>M. musculus</i> cathepsin L		<i>M. musculus</i> tubulin γ1		<i>M. musculus</i> average
FMC	1.380178	FMC	1.428569	CDH	1.439404	FMC	1.435972	FMC	1.437215	FMC	1.492206	FMC	1.406156	FMC	1.4074
CDH	1.393821	EMC	1.439538	BHC	1.440416	CDH	1.449731	EMC	1.454782	EMC	1.514578	EMC	1.422334	EMC	1.4202
EMC	1.393821	CDH	1.456844	EHC	1.443453	BHC	1.450991	CDH	1.456048	IMC	1.529776	BHC	1.437743	IMC	1.4411
BHC	1.396572	BHC	1.461218	MSC	1.445484	EMC	1.453517	BHC	1.458586	AMC	1.554584	CDH	1.437743	CDH	1.4486
EHC	1.396572	EHC	1.469392	FMC	1.446501	EHC	1.458586	EHC	1.46496	CDH	1.568813	EHC	1.456553	BHC	1.4548
Standard	1.404871	MSC	1.473186	Standard	1.450578	MSC	1.461131	MSC	1.46752	BHC	1.572003	MSC	1.459722	EHC	1.4596
MSC	1.404871	IMC	1.475089	EMC	1.458784	IMC	1.462405	IMC	1.470088	EHC	1.576807	Standard	1.463964	MSC	1.4653
AMC	1.410441	Standard	1.47509	IMC	1.462911	Standard	1.466239	Standard	1.472662	MSC	1.580022	IMC	1.463964	AMC	1.4687
IMC	1.418857	AMC	1.4853	AMC	1.469136	AMC	1.476535	AMC	1.484328	VMC	1.592988	AMC	1.485447	Standard	1.4706
VMC	1.44743	VMC	1.511292	VMC	1.479596	VMC	1.485632	VMC	1.496131	Standard	1.594621	VMC	1.507402	VMC	1.4961
AYHC	1.450333	YMC	1.569575	AYHC	1.528071	YMC	1.550345	YMC	1.558736	YMC	1.617763	YMC	1.570388	YMC	1.5612
YMC	1.513322	AYHC	1.57237	YMC	1.539169	AYHC	1.584343	AYHC	1.582903	AYHC	1.672759	AYHC	1.577506	AYHC	1.5799

	<i>F. catus</i> EF1 α		<i>F. catus</i> CD2		<i>F. catus</i> average
FMC	1.418236	FMC	1.445024	FMC	1.4038
EMC	1.430201	EMC	1.457692	EMC	1.4179
CDH	1.456626	CDH	1.48643	IMC	1.4444
IMC	1.457655	BHC	1.49816	CDH	1.4524
BHC	1.457656	EHC	1.504077	BHC	1.4577
EHC	1.458687	IMC	1.505562	EHC	1.4591
MSC	1.460753	MSC	1.513019	MSC	1.4649
Standard	1.465937	Standard	1.516018	AMC	1.4687
AMC	1.468018	AMC	1.53573	Standard	1.4716
VMC	1.484825	VMC	1.563682	VMC	1.4966
YMC	1.537011	YMC	1.589202	YMC	1.5591
AYHC	1.544854	AYHC	1.595685	AYHC	1.5929

When we have analyzed the effects of various coding tables on these genes as described in Methods section, we have observed that the Standard (or Universal) Code has performed significantly worse than many other coding tables, ranking between 6th and 9th among all 12 coding tables tested (Figure 2). In order to assess whether this low performance was simply due to a bias of these genes with respect to the average codon usage frequencies in the corresponding organisms, we have also generated so-called “average” genes, which are randomly representative of codon usage frequencies previously determined for those particular genomes (<http://www.kazusa.or.jp>) (Figure 2). The Standard Coding Table provided poor resilience towards mutations even in this hypothetical gene, whereas FMC performed much better in almost all of the human genes tested.

Of course, this could have been due to something peculiar about primate genes and requirements of these genomes. Thus, we have decided to analyze genes from representative mammals, essentially horse, cat and mouse (Figure 3). The ranking of the Standard Code among all other tables tested was pretty variable in mouse, albeit still low in terms of performance. Similar to primate genes, we wanted to check whether codon usage frequencies of these genomes could in fact have affected the analyses, and constructed hypothetical “average” genes also for these genomes. The Standard Code was still poor-performing, ranking 9th in all three organisms (Figure 3).

One of the better-performing nuclear coding tables in this analytical scheme was the CDH code, usually ranking between 2nd and 3rd positions – thus we wanted to address whether this nuclear code would still perform better when genes normally subject to this coding table were analyzed. To that end, we have used representative genes from 4 different species, all of which are involved in DNA replication machinery, and also constructed two “average” genes based on codon usage frequencies (Table 2). In all of the 4 genes selected, FMC was still a better-performing coding table, ranking the first in “survivability”, with one exception being the Hexamite Elongation Factor 1 (EF1) gene (Figure 3). Interestingly, Standard Code also performed slightly better for these genes, ranking between 4th and 6th among the 12 coding tables (Figure 3).

We have observed similar results in our analyses of ENC and MSC tables (data not shown): essentially, FMC was the best scoring table among all, with CDH in the top 4 in all of the assays (Figure 4).

The FMC appearing at the top of the list in most of the analyses was rather intriguing, especially since other mitochondrial coding tables have been usually the worst performing tables in all the analyses so far. Thus, we wanted to initially address the question of how this coding table performed with respect to the genes that normally utilizes this table, namely genes encoded by flatworm mitochondrial DNA. To that end, we have used several different mitochondrial genes from various flatworm species; in order to compare mitochondrial versus nuclear gene performance, and also constructed “average” genes for this table (Table 2). In all mitochondrial genes, FMC appeared as the highest-performing coding table, as with other species and coding tables examined so far (Figure 5). The Standard Code was still suboptimal for these genes, ranking 5th-8th among the 12 coding tables tested, while CDH is still a better performer than the Standard Code (Figure 5). For vertebrate or invertebrate mitochondrial tables, the genes encoded by mitochondrial genomes were analyzed, however the overall profile had not changed, with FMC still outperforming the Standard Code (data not shown).

Figure 4. Comparison of various coding tables with respect to robustness for selected genes that are normally subject to CDH Coding Table. “Average” genes represent hypothetical and idealized genes constructed using average codon usage frequencies (see Methods); accession numbers of the genes are listed in Table 2. Coding table abbreviations are given in Table 1.

	<i>Paraurostyla</i> DHA Pol		<i>Hexamita</i> EF1		<i>Urostyla</i> DHA Pol
FMC	1.5128	CDH	1.5053	FMC	1.5249
EMC	1.5387	EHC	1.5116	EMC	1.5474
CDH	1.5876	MSC	1.5129	CDH	1.5692
IMC	1.6019	Standard	1.5205	Standard	1.5894
Standard	1.6119	BHC	1.5205	EHC	1.5946
EHC	1.6205	FMC	1.5205	MSC	1.5991
MSC	1.6242	IMC	1.5333	IMC	1.6007
BHC	1.6246	EMC	1.5346	BHC	1.6032
AMC	1.6299	AMC	1.5372	AMC	1.6154
AYHC	1.6471	VMC	1.5411	VMC	1.6421
VMC	1.6721	AYHC	1.5555	AYHC	1.6472
YMC	1.7107	YMC	1.5728	YMC	1.6951

	<i>Tetrahymena</i> <i>thermophila</i> EF1		<i>Tetrahymena</i> <i>thermophila</i> EFG		<i>Tetrahymena</i> <i>thermophila</i> average
FMC	1.4445	FMC	1.5023	FMC	1.4892
CDH	1.4618	EMC	1.5203	EMC	1.5197
EHC	1.4717	CDH	1.5257	CDH	1.5756
EMC	1.4750	IMC	1.5325	Standard	1.5929
MSC	1.4917	EHC	1.5349	IMC	1.5978
Standard	1.4950	Standard	1.5398	EHC	1.6011
IMC	1.4984	MSC	1.5398	MSC	1.6078
BHC	1.4995	BHC	1.5454	BHC	1.6106
AMC	1.5282	AMC	1.5554	AYHC	1.6285
AYHC	1.5458	AYHC	1.5756	AMC	1.6302
VMC	1.5614	VMC	1.5775	VMC	1.6566
YMC	1.7649	YMC	1.5976	YMC	1.6958

	<i>Styloynchia</i> DHA Pol		<i>S. lemnae</i> micronuclear DHA pol		<i>S.lemnae</i> EF1		<i>S. lemnae</i> average
FMC	1.4038	FMC	1.4966	CDH	1.5108	FMC	1.4877
EMC	1.4145	EMC	1.5441	FMC	1.5119	EMC	1.5046
IMC	1.4809	EHC	1.6380	EMC	1.5199	CDH	1.5248
CDH	1.4847	IMC	1.6415	IMC	1.5210	IMC	1.5356
BHC	1.5002	CDH	1.6450	EHC	1.5221	EHC	1.5501
Standard	1.5041	Standard	1.6643	MSC	1.5221	Standard	1.5512
MSC	1.5080	BHC	1.6696	AMC	1.5244	BHC	1.5533
EHC	1.5119	MSC	1.6714	Standard	1.5290	MSC	1.5548
AMC	1.5278	AMC	1.6967	BHC	1.5290	AMC	1.5564
VMC	1.5770	AYHC	1.7004	VMC	1.5604	AYHC	1.5940
AYHC	1.5939	VMC	1.7474	AYHC	1.5640	VMC	1.5951
YMC	1.5939	YMC	1.8009	YMC	1.5652	YMC	1.6133

Figure 5. Comparison of various coding tables with respect to robustness for selected genes that are normally subject to FMC Coding Table. “Average” genes represent hypothetical and idealized genes constructed using average codon usage frequencies (see Methods); accession numbers of the genes are listed in Table 2. Coding table abbreviations are given in Table 1.

	<i>E. granulosus</i> IID1		<i>E. granulosus</i> malate dehydrogenase		<i>E. granulosus</i> thioredoxin glutathione reductase		<i>C. sinensis</i> cytochrome oxidase subunit 1
FMC	1.4514	FMC	1.3872	FMC	1.4268	FMC	1.3546
EMC	1.4977	CDH	1.3938	EMC	1.4456	EMC	1.3546
IMC	1.5266	EMC	1.3938	EHC	1.4728	CDH	1.4361
EHC	1.5464	EHC	1.3951	CDH	1.4744	EHC	1.4361
MSC	1.5699	BHC	1.4018	BHC	1.4768	Standard	1.4478
BHC	1.5801	Standard	1.4058	MSC	1.4833	BHC	1.4478
CDH	1.5801	MSC	1.4058	Standard	1.4858	MSC	1.4478
Standard	1.5836	IMC	1.4152	IMC	1.4883	IMC	1.4556
AMC	1.5905	AMC	1.4192	AMC	1.4941	AYHC	1.4795
AYHC	1.6331	VMC	1.4342	VMC	1.5049	AMC	1.4999
VMC	1.6367	YMC	1.4821	AYHC	1.5329	VMC	1.5123
YMC	1.7201	AYHC	1.4981	YMC	1.5546	YMC	1.5505

	<i>F. hepatica</i> cathepsin L		<i>T. asiatica</i> calcineurin B		<i>S. mansoni</i> CD2		<i>T. solium</i> dUTPase
FMC	1.4519	FMC	1.5159	FMC	1.4638	FMC	1.3576
EMC	1.4873	EMC	1.5248	EMC	1.4904	EMC	1.3722
IMC	1.5257	CDH	1.5584	CDH	1.5616	EHC	1.3870
AMC	1.5288	IMC	1.5709	EHC	1.5722	CDH	1.3930
CDH	1.5319	BHC	1.5741	MSC	1.5762	BHC	1.3960
EHC	1.5335	AMC	1.5804	BHC	1.5816	Standard	1.4051
MSC	1.5335	EHC	1.5836	Standard	1.5829	MSC	1.4051
BHC	1.5414	Standard	1.5868	AYHC	1.5951	IMC	1.4204
Standard	1.5462	MSC	1.5902	IMC	1.5978	AMC	1.4266
VMC	1.5574	VMC	1.6029	AMC	1.6144	VMC	1.4391
AYHC	1.5869	YMC	1.6460	VMC	1.6397	AYHC	1.4775
YMC	1.6241	AYHC	1.6631	YMC	1.6908	YMC	1.5244

To summarize all these scores in one table, we have taken the rankings of all coding tables used for each gene (Figure 6a). Afterwards, for all genes analyzed from that particular table (for instance MSC), these ranks were summed up and their averages were calculated (Figure 6b). When this was calculated for all the genes tested from all the different coding tables, we have organized our data in a tabulated form (Figure 6c). As can be seen, quite unexpectedly, Standard coding table ranged in performance ranking from 4.9 to 7.9, which was rather poor when compared to the FMC table, which was almost always 1st in the genes tested across species (Figure 3c). The worst performing table in almost all the cases was yet another mitochondrial table, YMC, which indicates that the results are not correlated with whether the genes are encoded by nuclear or mitochondrial genomes, or by mutations rates thereof. VMC was slightly better than YMC, but interestingly, IMC was closer to the Standard Code than to VMC (Figure 3c).

Figure 6. Ranks of coding tables across organisms and genes in terms of robustness and survivability. (A) A representative list of statistical calculations of survivability for a given gene if it were subject to different coding tables. (B) For genes that are normally decoded by one given Code Table, the average rank of certain alternate tables in terms of performance in survivability (projection of robustness). (C) A summary of performance comparison of all the coding tables tested with respect to one another. The rows show averages of all the genes analyzed for that particular coding scheme (ie, AVERAGE(FMC) means average ranks for genes that are normally subject to FMC), while the columns indicate what would have happened if alternative coding schemes were adopted.

A

Rank	Code	<i>T. rubrum</i> NADH4L
1	FMC	1.4138
	EMC	1.4343
	CDH	1.4930
4	Standard	1.4985
	BNC	1.5041
	ENC	1.5041
7	MSC	1.5041
	AYNC	1.5209
	IMC	1.5671
	AMC	1.5911
	VMC	1.6281
12	YMC	1.7272

B

preferred code	gene name	FMC rank	STD rank	YMC rank		MSC rank
MSC	<i>T. rubrum</i> NADH4L	1	4	12		7
MSC	<i>T. rubrum</i> cox 2	2	3	12		7
MSC	<i>T. rubrum</i> cox1	1	5	12		7
MSC	<i>T. rubrum</i> average	1	8	12		6
MSC	<i>N. crassa</i> cox 2	1	6	12		8
MSC	<i>N. crassa</i> NADH	1	6	12		8
MSC	<i>N. crassa</i> average	1	8	12		6
AVERAGE (MSC)		1.1	5.7	12.0		7.0

C

	FMC	STD	YMC	CDH	ENC	MSC	VMC	IMC
AVERAGE (FMC)	1.0	7,1	11,8	4,8	4,2	6,3	10,3	5,7
AVERAGE (STD)	1,1	7,9	11,3	3,3	5,4	6,7	10,0	5,6
AVERAGE (YMC)	1,0	4,9	12,0	3,9	3,4	7,3	11,0	9,1
AVERAGE (CDH)	1,5	5,5	12,0	2,8	4,8	6,4	10,6	4,8
AVERAGE (ENC)	1,0	8,5	11,8	3,8	5,7	6,9	10,6	4,5
AVERAGE (MSC)	1,1	5,7	12,0	4,1	5,1	7,0	10,7	5,6
AVERAGE (VMC)	1,0	6,4	11,6	4,3	4,7	6,5	10,2	6,1
AVERAGE (IMC)	1,2	5,2	11,8	5,3	4,0	5,7	10,8	8,1

3. Conclusions

In this paper, we have used statistical analysis to investigate the optimality of alternate genetic coding tables on the “survivability” of genes representative of different organisms. This analysis simply calculates the probability of maintenance of a functional protein after generations of single-base substitutions in a given gene, depending on whether the mutation is silent, missense or nonsense in that particular coding table used (see Methods). Our results indicate that the Standard, or “Universal”, Genetic Code is actually one of the lower performance tables in terms of tolerating mutations and rendering another functional protein upon genetic substitutions. The best success rates were obtained, surprisingly, with FMC for all the organisms tested, which was not paralleled in either the VMC or other mitochondrial coding tables analyzed (Figure 6).

Interestingly, the CDH does in fact give the maximum score for the *Hexamita* gene, EF1, which in fact does use this very coding table, and CDH performs significantly better than the Standard Code for other ciliate and *Hexamita* genes that we have analyzed. This may imply that indeed CDH might be evolutionarily more adapted to the organism and the environment against any possible mutations (see below).

There have been many studies trying to explain the presence of alternate genetic codes, or indeed why the standard genetic code is still undergoing reassignments of codons [5,7,9,10], which required modification of the so-called “frozen-accident” theory [1]. These changes in alternate code tables are believed to have stemmed from reassignments of codons of the Standard Genetic Code, and not from ancestral lineages of alternate coding tables [10]. Then one could imagine a situation where newly evolved codes due to evolution may indeed be better suited for certain organisms and certain conditions, in line with our data on FMC, nevertheless this still fails to explain why FMC and not any other mitochondrial code?

This could in part be explained by the nature of codon reassignments in these particular code tables: when IMC, VMC, YMC, and FMC are all compared, there are a few common reassignment schemes – UGA that is STOP in the Standard Code is reassigned to Trp in these mitochondrial codes, and AUA that is Ile in the Standard code is assigned to Met in all these codes except for FMC (see Table 1 and Figure 1). However, when the differential reassignments are analyzed, it becomes apparent that there is an entire subset of codons for Leucine (CUU, CUC, CUA and CUG) that has been reassigned to another amino acid, Threonine (Table 1) in YMC table, leaving only two codons still encoding Leu. Also in YMC, the UGA STOP codon has been reassigned to Trp, which would lead to failure to stop translation of certain proteins. Furthermore, to make things even worse for YMC, two codons, CGC and CGA, have been left unassigned (Table 2). Similarly, in the VMC scheme, AGA and AGG codons for Arginine have been reassigned to STOP, which could result in immature termination of translation, thus affecting the performance of this coding table immensely.

When the FMC table was analyzed, however, one can readily observe that the reassignments are relatively “mild” when compared to the other coding tables: the UAA STOP codon has been changed to Tyrosine, reducing one STOP codon, however at the same time increasing the robustness to any mutations to the two Tyrosine codons, UAU and UAC (see Figure 1 and Table 2). The AAA codon for the positively-charged Lys has been reassigned to a polar Asn residue, and the two Arg codons, namely AGA and AGG, have been reassigned to another polar residue, Ser (Table 2). When compared to our results,

these changes would appear to have increased the tolerance of this coding table to mutations that affect primary structure of the protein (Figure 5). Of course, one must note that this initial study excludes any similarity matrices for mutations as it is, however recent data indicate that improvement of the calculations based on input from BLOSUM, PAM (Point Accepted Mutations) and other matrices does not significantly alter the ranking profile of these coding tables (Kurnaz and Kurnaz, unpublished data). This could partly be explained by the fact that PAM matrix itself may be a result of the nature of the genetic code table [11].

When mechanistics of evolution of alternate genetic codes are investigated, it appears that almost all of the present-day alternate codes are much “better” than any random genetic code table constructed [4], although in their analyses of how reassignments can have occurred the researchers conclude that the “canonical” or standard genetic code is slightly better than the alternates. If one assumes error-minimization as the basic premise of optimality of code tables, then the standard genetic code appears to be the most optimized, as well as most adaptive when compared to alternative genetic codes [11]. In our study we observe that the Standard code is not the best among the alternate codes tested, in terms of robustness and toleration of mutations. This is probably due to our calculations taking into account only robustness of the genetic codes – however one study indicates that changeability of sequences is just as important for the evolution of genetic codes, thus adaptability, where the alternate genetic codes would suffer in our calculations [11, 12]: although seemingly contradictory, in this work robustness and changeability was implied to be equally fundamental for the survivability of organisms and evolution of genetic codes. Changeability is defined as a measure of how much a sequence can be altered through single base mutations [12], which, while putting a certain population of organisms at a certain disadvantage, could also lead to a slight advantage in another subset of organisms. This would be one explanation as to why the Standard Code, although so poor in terms of robustness, is the most widely-used coding table of all. As other researchers point out, although contemporary genomes operate in almost error-free environments, ancestors of the standard code were most likely in a highly error-prone niche, where robustness would have held a certain disadvantage [11]. A different explanation, in the light of still ongoing codon reassignments, could equally well be that the genetic codes are still changing and even possibly, evolving. Our hypothesis is more in line with the latter explanation, where the standard code was quite possibly the first optimal scheme reached by natural selection, but that it is still evolving, both reassigning certain codons, as well as expanding the amount of information contained by the table to the optimal number calculated for this coding table, 22 [5,7,8,9,10]. However such hypotheses need to be further tested after additional parameters such as the changeability have been included in the calculations. Also, combinatorics approaches could be utilized in order to provide comparison to the results obtained through the calculations presented in this study.

4. Methods

If there is a mutation on a gene which causes a severe change in the amino acid chain, we could safely assume that the organism would not be able to build a functional protein. If this protein is a crucial protein for the viability of the organism, one could also assume that any deleterious mutations in this gene rendering the protein non-functional would be lethal to the organism. Proteins taking part in immunity, cell respiration, DNA, RNA synthesis and cell division (tubulin formation) are basic examples

for such crucial proteins. This has been previously reported to be the case for a human cytokine gene used in simulations and statistical calculations, following experimental findings in literature [8]. If all other effects (aging, food restriction, illness etc.) that are not directly related to the genetic code table are neglected, deleterious mutations changing the amino acid sequence will be the major cause of death in this *in silico* population. Neutral or “silent” mutations do not cause a change in viability in this model (see probability calculations below). We have also omitted reproduction from this model for simplicity, therefore we have a population which can only decrease as a result of deleterious mutations in order to emphasize the effects of coding tables alone.

There are many different schemes on the occurrence of mutations – in this preliminary model we have assumed that a mutation is a random process, with equal probability [13]. We disregard frameshift mutations in this particular model (caused by deletions or insertions), and we only look at single nucleotide substitutions – hence we include in this model the effects of only silent mutations, nonsense mutations and missense mutations on the protein product. Normally the rates for these replacements depend on the two nucleotides being interchanged. The simplest approach to the problem is to take all mutation rates to be equal, an approach known as the Jukes-Cantor mutation scheme [14].

The mutation is taken to be deleterious if it causes a change in the amino acid chain; and not all the mutations kill the individual. To be more explicit, the codons AAA and AAG in the Standard Genetic code the same amino acid, “lysine”; hence if AAA turns into AAG as a result of a mutation the amino acid will not change and the protein can be constructed safely. However; if AAA turns into AGA, which codes the amino acid “arginine”, the amino acid chain will change and we assume that the protein can not build up, which means the represented organism will die.

There can be a mutation which converts AAA to AAX where $X \neq \{A, G, C, \text{ or } T\}$; then the individual dies automatically. As a model, we are looking at a simpler case where a mutation changes A to one of G, C, or T, but not X. Since reproduction is not included in the model, the population can only diminish. The decrease in population can be found by calculating the probability of a deleterious mutation. The details of these calculations can be found in [8]. Essentially, the probability of the mutation changing the amino acid depends on the codon; so one needs to find the probability of hitting each different codon type. First, the probability of hitting a codon type (P_α) is calculated as the ratio of the number of codons of that type in the gene (N_α) to total number of codons. Then we need to exclude the mutations that do not cause a change in the amino acid and calculate the probability of a change occurring in the amino acid caused by a change in one nucleotide ($P(d/\alpha)$). The results are reported as the negative of the slope, hence the smaller numbers indicate better survival rates over many generations. In general, we let the populations continue over a minimum of 10 generations.

We used only the exon (protein coding) part of the gene considering any mutation in the intron would be essentially harmless with respect to amino acid substitutions. As a simple example, the human cytokine gene has a total length of 2068 nucleotides; 621 nucleotides in exon part and 1447 ones in intron. The probability of hitting the exon part of the gene is simply the ratio of the exon part to the total gene:

$$P(\text{hitting exon}) = \frac{621}{2068} = 0.3032 \quad (1)$$

Hence; the probability of having a *deleterious* mutation for all the gene is simply a product of mutation probability and probability of hitting the exon part of gene. As the chances of hitting any part of the gene is a same, we can neglect the intron part in the simulation since this would only be a multiplicative constant in the problem. Therefore the probability of having a deleterious mutation for all human cytokine gene is simply:

$$P(\text{deleterious}) \propto \sum_{\alpha=1}^{64} P_{\alpha} P(d/\alpha) = 0.7729, \quad (2)$$

where d is the number of deleterious mutations and α is the number of all possible mutations. Therefore, for our purposes, we have not used genomic sequences but rather CDS, or coding sequences, for the sake of simplicity because of the calculations discussed above.

The survival probability can be calculated by:

$$P(\text{surviving}) = 1 - P(\text{deleterious}) = 0.2271 \quad (3)$$

If we take an initial population of N_0 genes (individuals), after n number of mutations, to the first order, the number of surviving individuals (N_n) is given by:

$$N_n \approx N_0 P(\text{surviving})^n \quad (4)$$

Hence, we obtain the “probability of survival” with the slope of the number of surviving individuals versus time graph:

$$\text{slope} \approx \ln[P(\text{surviving})] = -1.4823 \quad (5)$$

Similarly the probability of survival can be calculated for all the genes separately. However in this calculation once we make a change in the gene sequence and if the individual survives, we forget about the change we have made and restart the process for the second mutation cycle with the original gene sequence. We have assumed that in Nature, if the individual survives, the second mutation cycle starts with the mutated gene sequence and not the original one. Therefore, to be able to get closer to Nature we have also written a simulation code which allows for the mutation in the gene sequence to be kept in the next mutation mutation cycle.

The Standard Genetic Coding Table is shown in Figure 1. The variations of the different Coding Tables when compared to the Standard Coding Table are summarized in Table 1. The genes that have been used for this study are summarized in Table 2.

Acknowledgements

We would like to acknowledge Ms. Evin Gultepe, who has generated the original simulation program that had preceded this statistical study. We would also like to thank Dr. Muhittin Mungan for helpful discussions.

References and Notes

1. Crick, F.H.C. The origin of the genetic code. *J. Mol. Biol.* **1968**, *38*, 367-379.
2. Sella, G.; Ardell, D.H. The Coevolution of genes and genetic codes: Crick's frozen accident revisited. *J. Mol. Evol.* **2006**, *63*, 297-313.
3. Sengupta, S.; Yang, X.; Higgs, P.G. The mechanisms of codon reassignments in mitochondrial genetic codes. *J. Mol. Evol.* **2007**, *64*, 662-688.
4. Sengupta, S.; Higgs, P.G. A unified model of codon reassignment in alternative genetic codes. *Genetics* **2005**, *170*, 831-840.
5. Ardell, D.H. On error minimization in a sequential origin of the standard genetic code. *J. Mol. Evol.* **1998**, *47*, 1-13.
6. Wilke, C.O.; Adami, C. Evolution of mutational robustness. *Mut. Res.* **2003**, *522*, 3-11.
7. Haig, D.; Hurst, L.D. A quantitative measure of error minimization in the genetic code. *J. Mol. Evol.* **1991**, *33*, 412-417.
8. Gultepe, E.; Kurnaz, M.L. Monte Carlo simulation and statistical analysis of genetic information coding. *Physica A* **2005**, *357*, 525-533.
9. Telford, M.J.; Herniou, E.A.; Russell, R.B.; Littlewood, D.T.J. Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. *Proc. Natl. Acad. Sci.* **2000**, *97*, 11359-11364.
10. Silva, R.M.; Miranda, I.; Moura, G.; Santos, M.A. Yeast as a model organism for studying the evolution of non-standard genetic codes. *Genomics and Proteomics* **2004**, *3*(1), 35-46.
11. Freeland, S.J.; Knight, R.D.; Landweber, L.F.; Hurst, L.D. Early fixation of an optimal genetic code. *Mol. Biol. Evol.* **2000**, *17*(4), 511-518.
12. Maeshiro, T.; Kimura, M. The role of robustness and changeability on the origin and evolution of genetic codes. *Proc. Natl. Acad. Sci.* **1998**, *95*, 5088 – 5093.
13. Volkenshtein, M.V. Probabilities of transversions and transitions. *Mol. Biol. (Mosk.)* **1976**, *10*(4), 605- 608.
14. Jukes, T.H.; Cantor, J.R. Evolution of protein molecules. In *Mammalian Protein Metabolism*; Munro, H.A., Ed.; Academic Press: New York, NY, USA, 1969; pp. 21-132.

© 2008 by the authors; licensee Molecular Diversity Preservation International, Basel, Switzerland. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).