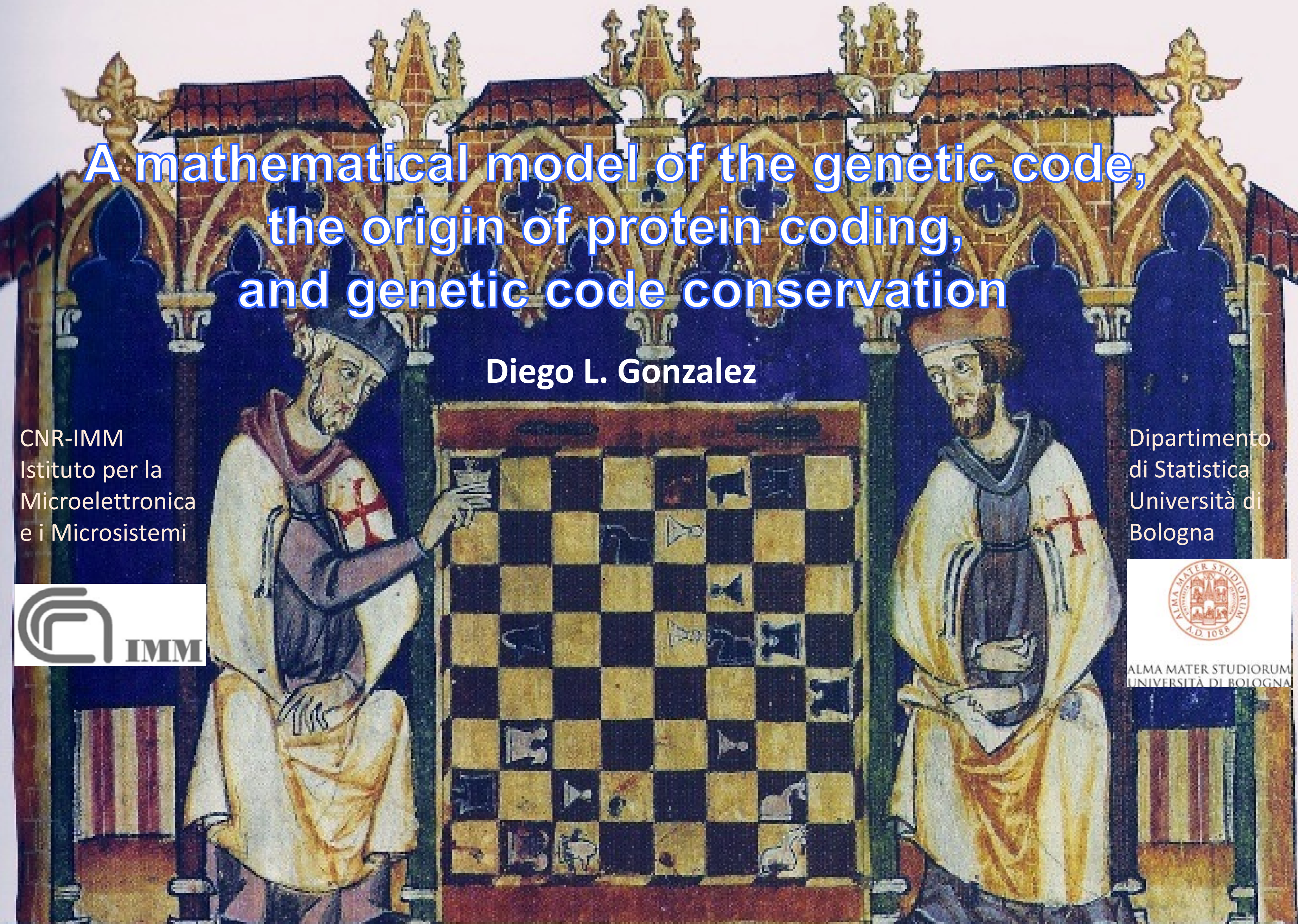


A mathematical model of the genetic code, the origin of protein coding, and genetic code conservation

Diego L. Gonzalez

CNR-IMM
Istituto per la
Microelettronica
e i Microsistemi

Dipartimento
di Statistica
Università di
Bologna

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Biological Codes

- **Biological codes** are *the great* ***invariants*** *of life*, the sole entities that have been **conserved** in evolution while everything else has changed, *Barbieri*, 2014.

- What has been conserved in biological codes?

- A**daptors**, in short, are the **key** molecules in all **organic codes**. They are the molecular fingerprints of the codes. … It is one of those facts that have extraordinary **theoretical** implications

- If adaptors are the key molecules, the origin of the codes should be described by the properties of the adaptors?

- In biology, conserved quantities are related to relevant biological functions; in physics conserved quantities are a consequence of **symmetry**. In a physical paradigm, symmetry is our first ingredient in the search for conserved quantities in biological codes.

Degeneracy

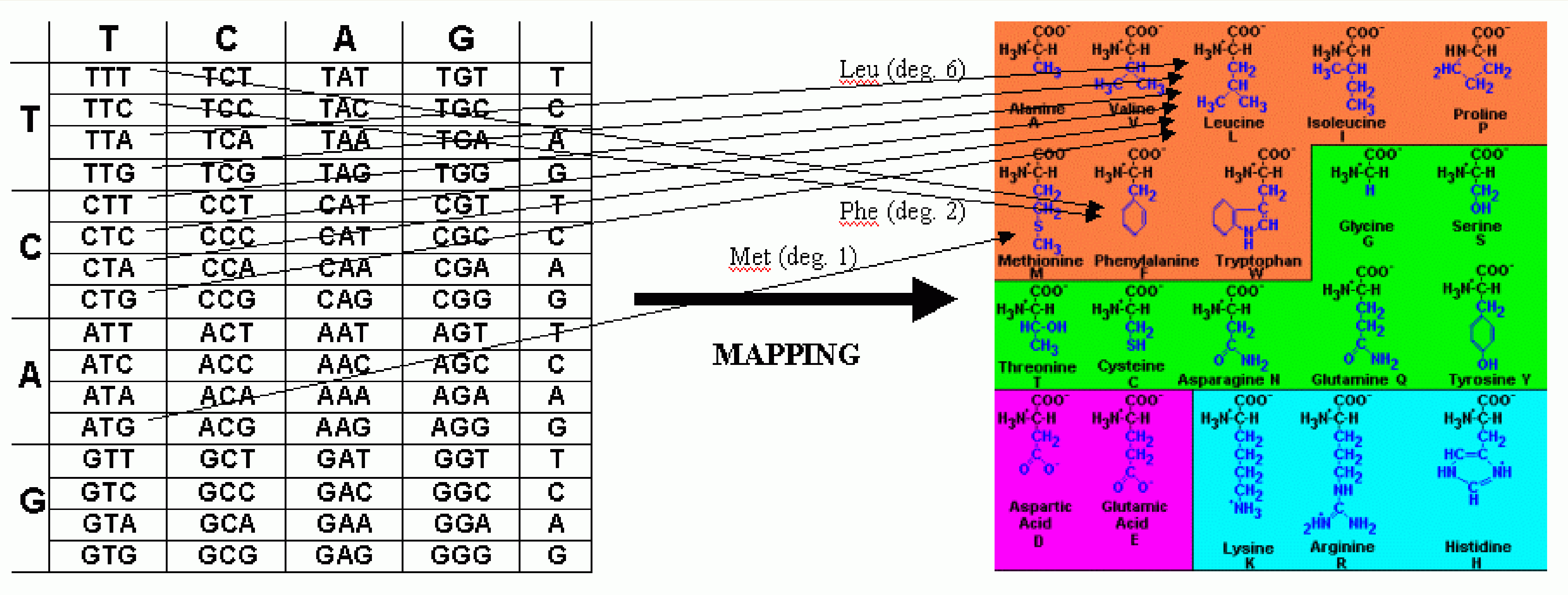
- Focusing on the genetic code: is it possible to construct a model of the **genetic code** based on **adaptors** and through this model is it possible to study **conserved** quantities?

- **What is degeneracy?**

Degeneracy is a concept that comes from quantum mechanics. It refers to the fact that the same total energy can characterize more that one different quantum-state. By analogy we say that an amino acid is degenerate because it is associated to more than one different codon.

- **Degeneracy is the main mathematical property of the genetic code**

The Genetic Code as a Mapping



64 codons; **4x4x4 = 64** (Words of three letters from an alphabet of four, i.e., A, T, C, G)

20 amino acids + Stop codon (Methyonine represents also the synthesis start signal)

→ **Redundancy and Degeneracy follow**

Degeneracy Distribution of the Nuclear Euplotes Genetic Code

Second letter

First letter	U	C	A	G	Third letter
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
U	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
U	UUA Leu	UCA Ser	UAA Stop	UGA Cys	A
U	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
C	CUC Leu	CCC Pro	CAC His	CGC Arg	C
C	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
C	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
A	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
A	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
A	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
G	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
G	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
G	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

Euplotes nuclear genetic code

Degeneracy distribution inside quartets

Degeneracy	# of amino acids
4	8
3	2
2	12
1	2

Mathematical Modelling of the Genetic Code

•In the last years we have developed a mathematical model of the genetic code that explains the degeneracy distribution of both variants of the genetic code, nuclear and mitochondrial.

•The model is based on number representation systems. Numbers are usually represented in an univocal way: any number has only one representation and any representation corresponds to only one number.

•This is the case, for example, of the common decimal representation system. This univocity is due to the characteristics of the positional values and digits used in the so called power numeration systems (because the positional values of the representation correspond to the powers of a given base, i.e., 10 in the decimal system).

•**As the genetic code is degenerate and redundant we are interested in non-univocal representation systems**

Mathematical model: integer number representation systems

Usual positional notation – power representation systems – base k

d_5	d_4	d_3	d_2	d_1	d_0
k^5	k^4	k^3	k^2	k^1	k^0

Digits: d_i (0, (k-1))

$$R_{\text{(integer)}} = d_5k^5 + d_4k^4 + d_3k^3 + d_2k^2 + d_1k^1 + d_0k^0$$

Univocity condition: $$(k-1)\sum_{n=0}^{m} k^n = k^{m+1} - 1$$

Example 1: k=10; decimal system

Digits: d_i (0, 9)

0	0	0	0	1	9
10^5	10^4	10^3	10^2	10^1	10^0

$$R = 1.10^1 + 9.10^0 = 19$$

Redundant representation systems

Example 2: Binary power system

0	1	0	0	1	1
2^5	2^4	2^3	2^2	2^1	2^0

Digits: d_i (0, 1)

$$R = 0.2^5 + 1.2^4 + 0.2^3 + 0.2^2 + 1.2^1 + 1.2^0 = 19$$

Redundant Power Representations Using Out-of-Range Digits

Example 3: Binary signed digits with three possible values: **-1, 0, 1**

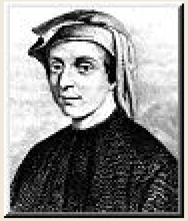
1	-1	0	0	1	1
2^5	2^4	2^3	2^2	2^1	2^0

$$R = 1.2^5 + 0.2^4 - 1.2^3 + 0.2^2 + 1.2^1 - 1.2^0 = 19$$ **Redundant**

Non-power redundant representation systems

positional bases grow slowly than the powers of k

Example 4: k=2; Fibonacci system



1	1	1	1	0	1
8	5	3	2	1	1

Digits: d_i (0, 1)

R = 1.8 + 1.5 + 1.3 + 1.2 + 0.1 + 1.1 = 19

Non-power representation systems are redundant, for example, the number 19 can be represented in the Fibonacci system in another alternative way:

1	1	1	1	1	0
8	5	3	2	1	1

R = 1.8 + 1.5 + 1.3 + 1.2 + 1.1 + 0.1 = 19

Maya serpent numbers

Fibonacci's non-power representation system

	Binary Strings																							
Bases	8	5	3	2	1	1	8	5	3	2	1	1	8	5	3	2	1	1	8	5	3	2	1	1
Number																								
0	0	0	0	0	0	0																		
1	0	0	0	0	0	1	0	0	0	0	1	0												
2	0	0	0	1	0	0	0	0	0	0	1	1												
3	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	1	0	1						
4	0	0	1	0	1	0	0	0	1	0	0	1	0	0	0	1	1	1						
5	0	1	0	0	0	0	0	0	1	1	0	0	0	0	1	0	1	1						
6	0	1	0	0	1	0	0	1	0	0	0	1	0	0	1	1	1	0	0	0	1	1	0	1
7	0	1	0	1	0	0	0	1	0	0	1	1	0	0	1	1	1	1						
8	1	0	0	0	0	0	0	1	1	0	0	0	0	1	0	1	1	0	0	1	0	1	0	1
9	1	0	0	0	1	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1	1	0	0	1	0	1	0	1	1	1
10	1	0	0	1	0	0	1	0	0	0	1	1	0	1	1	1	0	0	0	1	1	0	1	1
11	0	1	1	1	0	1	0	1	1	1	1	0	1	0	0	1	0	1	1	0	0	1	1	0	1	0	1	0	0	0
12	0	1	1	1	1	1	1	0	0	1	1	1	1	0	1	0	0	1	1	0	1	0	1	0
13	1	0	1	0	1	1	1	0	1	1	0	0	1	1	0	0	0	0						
14	1	0	1	1	0	1	1	0	1	1	1	0	1	1	0	0	0	1	1	1	0	0	1	0
15	1	0	1	1	1	1	1	1	0	0	1	1	1	1	0	1	0	0						
16	1	1	0	1	0	1	1	1	0	1	1	0	1	1	1	0	0	0						
17	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	0	1	0						
18	1	1	1	0	1	1	1	1	1	1	0	0												
19	1	1	1	1	1	0	1	1	1	1	0	1												
20	1	1	1	1	1	1																		

Is there a redundant representation describing the degeneracy distribution of the genetic code?

Degeneracy	Number
Signed Binary	
1	5
2	3
3	4
4	2
5	2
Fibonacci System	
1	2
2	4
3	8
4	5
5	2
Genetic Code	
1	2
2	12
3	2
4	8

Non-power representation system 1,1,2,4,7,8

Represented number	Length 6 binary strings																							
	8	7	4	2	1	1	8	7	4	2	1	1	8	7	4	2	1	1	8	7	4	2	1	1
0	0	0	0	0	0	0																		
1	0	0	0	0	0	1	0	0	0	0	1	0												
2	0	0	0	0	1	1	0	0	0	1	0	0												
3	0	0	0	1	0	1	0	0	0	1	1	0												
4	0	0	1	0	0	0	0	0	0	1	1	1												
5	0	0	1	0	0	1	0	0	1	0	1	0												
6	0	0	1	1	0	0	0	0	1	0	1	1												
7	0	0	1	1	0	1	0	0	1	1	1	0	0	1	0	0	0	0						
8	0	1	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	1	1	1
9	1	0	0	0	0	1	1	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0	1	1
10	0	1	0	1	0	1	0	1	0	1	1	0	1	0	0	1	0	0	1	0	0	0	1	1
11	1	0	0	1	0	1	1	0	0	1	1	0	0	1	1	0	0	0	0	1	0	1	1	1
12	0	1	1	0	0	1	0	1	1	0	1	0	1	0	1	0	0	0	1	0	0	1	1	1
13	1	0	1	0	0	1	1	0	1	0	1	0	0	1	1	1	0	0	0	1	1	0	1	1
14	0	1	1	1	0	1	0	1	1	1	1	0	1	0	1	1	0	0	1	0	1	0	1	1
15	1	0	1	1	0	1	1	0	1	1	1	0	1	1	0	0	0	0	0	1	1	1	1	1
16	1	1	0	0	0	1	1	1	0	0	1	0	1	0	1	1	1	1						
17	1	1	0	0	1	1	1	1	0	1	0	0												
18	1	1	0	1	0	1	1	1	0	1	1	0												
19	1	1	1	0	1	1	1	1	0	1	1	1												
20	1	1	1	0	0	1	1	1	1	0	1	0												
21	1	1	1	1	0	0	1	1	1	0	1	1												
22	1	1	1	1	0	1	1	1	1	1	1	0												
23	1	1	1	1	1	1																		

Degeneracy distribution Non-power representation system 1,1,2,4,7,8	
# of integer numbers	Degeneracy
2	1
12	2
2	3
8	4

Unique solution 1,1,2,4,7,8

Degeneracy distribution Euplotes nuclear genetic code	
# of amino acids	Degeneracy
2	1
12	2
2	3
8	4

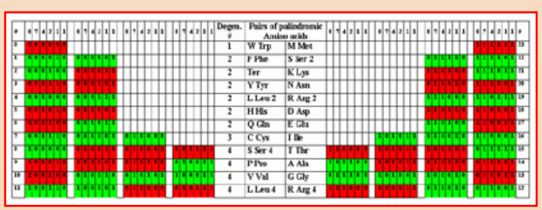
From a structural isomorphism to a true model

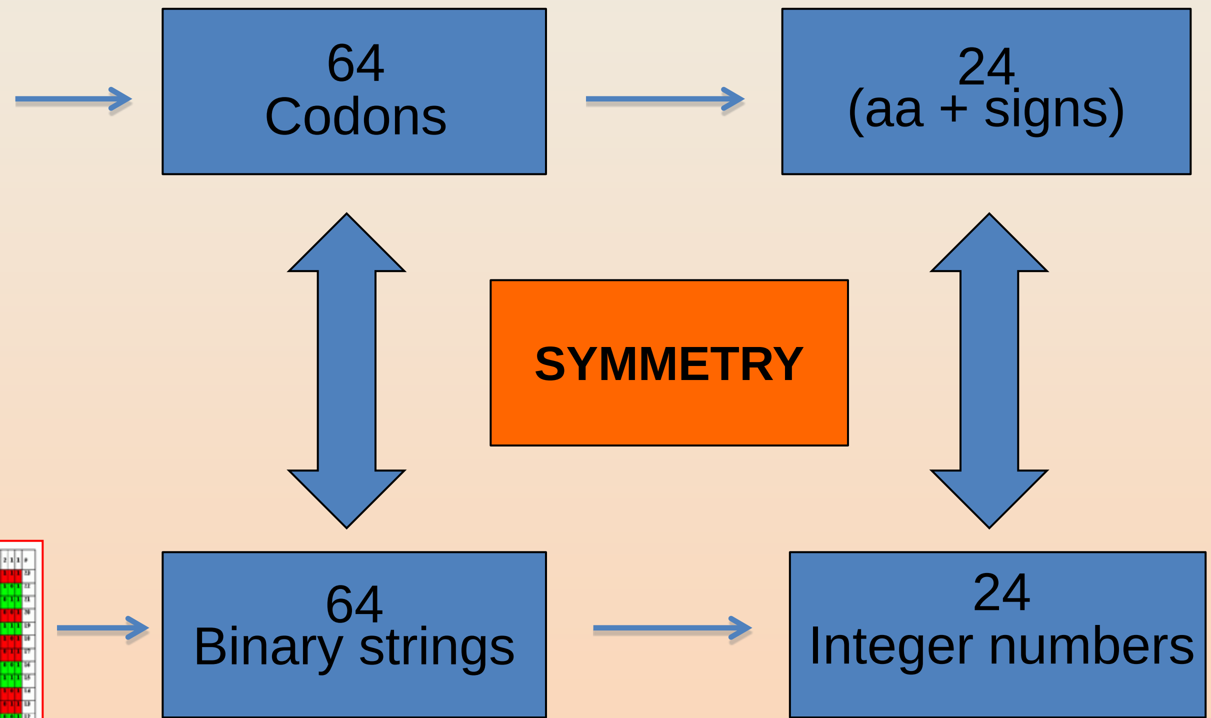
GENETIC CODE

Second letter

First letter	U	C	A	G	Third letter
U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA Stop UAG Stop	UGU Cys UGC Cys UGA Stop UGG Trp	U C A G
C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg	U C A G
A	AUU Ile AUC Ile AUA Ile AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg	U C A G
G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gly GGC Gly GGA Gly GGG Gly	U C A G

NON-POWER REPRESENTATION





The Model of the Euplotes Nuclear Genetic Code

	U			C			A			G			
U	1	000001	Phe	15	101101	Ser	18	110110	Tyr	16	110010	Cys	U
	1	000010	Phe	15	101110	Ser	18	110101	Tyr	16	110001	Cys	C
	4	001000	Leu	15	011111	Ser	2	000100	Ter	16	101111	Cys	A
	11	011000	Leu	15	110000	Ser	2	000011	Ter	23	111111	Trp	G
C	11	100101	Leu	14	011110	Pro	3	000101	His	12	011010	Arg	U
	11	100110	Leu	14	011101	Pro	3	000110	His	12	011001	Arg	C
	4	000111	Leu	14	101100	Pro	17	110100	Gln	19	111000	Arg	A
	11	010111	Leu	14	101011	Pro	17	110011	Gln	12	101000	Arg	G
A	7	001101	Ile	8	010010	Thr	5	001001	Asn	22	111110	Ser	U
	7	001110	Ile	8	010001	Thr	5	001010	Asn	22	111101	Ser	C
	7	010000	Ile	8	100000	Thr	21	111011	Lys	19	110111	Arg	A
	0	000000	Met	8	001111	Thr	21	111100	Lys	12	100111	Arg	G
G	13	101001	Val	9	100001	Ala	20	111010	Asp	10	010110	Gly	U
	13	101010	Val	9	100010	Ala	20	111001	Asp	10	010101	Gly	C
	13	011100	Val	9	010011	Ala	6	001011	Glu	10	100011	Gly	A
	13	011011	Val	9	010100	Ala	6	001100	Glu	10	100100	Gly	G

"Can the genetic code be mathematically described?", Diego L. Gonzalez, *Medical Science Monitor, v.10, n.4, HY11-17 (2004)*.
"The Mathematical Structure of the Genetic Code" D.L. Gonzalez, *The Codes of Life, M.. Barbieri Editor, Springer Verlag (2008)*

Symmetry of U ←→ C exchange

NNU and NNC codify the same amino acid

16 pairs of codons NNU - NNC

	U		C		A		G		
U	Phe	UUU	Ser	UCU	Tyr	UAU	Cys	UGU	U
		UUC		UCC		UAC		UGC	C
	Leu	UUA		UCA	Stop	UAA		UGA	A
		UUG		UCG		UAG	Trp	UGG	G
C	Leu	CUU	Pro	CCU	His	CAU	Arg	CGU	U
		CUC		CCC		CAC		CGC	C
		CUA		CCA	Gln	CAA		CGA	A
		CUG		CCG		CAG		CGG	G
A	Ile	AUU	Thr	ACU	Asn	AAU	Ser	AGU	U
		AUC		ACC		AAC		AGC	C
		AUA		ACA	Lys	AAA	Arg	AGA	A
	Met	AUG		ACG		AAG		AGG	G
G	Val	GUU	Ala	GCU	Asp	GAU	Gly	GGU	U
		GUC		GCC		GAC		GGC	C
		GUA		GCA	Glu	GAA		GGA	A
		GUG		GCG		GAG		GGG	G

Common to all known versions of the genetic code

Symmetry of U ←→ C exchange

NNU and NNC codify the same amino acid

16 pairs of codons NNU - NNC

	U		C		A		G		
U	Phe	UUU	Ser	UCU	Tyr	UAU	Cys	UGU	U
		UUC		UCC		UAC		UGC	C
	Leu	UUA		UCA	Stop	UAA		UGA	A
		UUG		UCG		UAG	Trp	UGG	G
C	Leu	CUU	Pro	CCU	His	CAU	Arg	CGU	U
		CUC		CCC		CAC		CGC	C
		CUA		CCA	Gln	CAA		CGA	A
		CUG		CCG		CAG		CGG	G
A	Ile	AUU	Thr	ACU	Asn	AAU	Ser	AGU	U
		AUC		ACC		AAC		AGC	C
		AUA		ACA	Lys	AAA	Arg	AGA	A
	Met	AUG		ACG		AAG		AGG	G
G	Val	GUU	Ala	GCU	Asp	GAU	Gly	GGU	U
		GUC		GCC		GAC		GGC	C
		GUA		GCA	Glu	GAA		GGA	A
		GUG		GCG		GAG		GGG	G

Common to all known versions of the genetic code

Non-power representation system 1,1,2,4,7,8

Represented number	Length 6 binary strings																							
	8	7	4	2	1	1	8	7	4	2	1	1	8	7	4	2	1	1	8	7	4	2	1	1
0	0	0	0	0	0	0																		
1	0	0	0	0	0	1	0	0	0	0	1	0												
2	0	0	0	0	1	1	0	0	0	1	0	0												
3	0	0	0	1	0	1	0	0	0	1	1	0												
4	0	0	1	0	0	0	0	0	0	1	1	1												
5	0	0	1	0	0	1	0	0	1	0	1	0												
6	0	0	1	1	0	0	0	0	1	0	1	1												
7	0	0	1	1	0	1	0	0	1	1	1	0	0	1	0	0	0	0						
8	0	1	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	1	1	1
9	1	0	0	0	0	1	1	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0	1	1
10	0	1	0	1	0	1	0	1	0	1	1	0	1	0	0	1	0	0	1	0	0	0	1	1
11	1	0	0	1	0	1	1	0	0	1	1	0	0	1	1	0	0	0	0	1	0	1	1	1
12	0	1	1	0	0	1	0	1	1	0	1	0	1	0	1	0	0	0	1	0	0	1	1	1
13	1	0	1	0	0	1	1	0	1	0	1	0	0	1	1	1	0	0	0	1	1	0	1	1
14	0	1	1	1	0	1	0	1	1	1	1	0	1	0	1	1	0	0	1	0	1	0	1	1
15	1	0	1	1	0	1	1	0	1	1	1	0	1	1	0	0	0	0	0	1	1	1	1	1
16	1	1	0	0	0	1	1	1	0	0	1	0	1	0	1	1	1	1						
17	1	1	0	0	1	1	1	1	0	1	0	0												
18	1	1	0	1	0	1	1	1	0	1	1	0												
19	1	1	1	0	1	1	1	1	0	1	1	1												
20	1	1	1	0	0	1	1	1	1	0	1	0												
21	1	1	1	1	0	0	1	1	1	0	1	1												
22	1	1	1	1	0	1	1	1	1	1	1	0												
23	1	1	1	1	1	1																		

Degeneracy distribution Non-power representation system 1,1,2,4,7,8	
# of integer numbers	Degeneracy
2	1
12	2
2	3
8	4

Unique solution 1,1,2,4,7,8

Degeneracy distribution Euplotes nuclear genetic code	
# of amino acids	Degeneracy
2	1
12	2
2	3
8	4

Symmetry of xxxx01 ←→ xxxx10 exchange

Represented number	Length 6 binary strings																							
	8	7	4	2	1	1	8	7	4	2	1	1	8	7	4	2	1	1	8	7	4	2	1	1
0	0	0	0	0	0	0																		
1	0	0	0	0	0	1	0	0	0	0	1	0												
2	0	0	0	0	1	1	0	0	0	1	0	0												
3	0	0	0	1	0	1	0	0	0	1	1	0												
4	0	0	1	0	0	0	0	0	0	1	1	1												
5	0	0	1	0	0	1	0	0	1	0	1	0												
6	0	0	1	1	0	0	0	0	1	0	1	1												
7	0	0	1	1	0	1	0	0	1	1	1	0	0	1	0	0	0	0						
8	0	1	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	1	1	1
9	1	0	0	0	0	1	1	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0	1	1
10	0	1	0	1	0	1	0	1	0	1	1	0	1	0	0	1	0	0	1	0	0	0	1	1
11	1	0	0	1	0	1	1	0	0	1	1	0	0	1	1	0	0	0	0	1	0	1	1	1
12	0	1	1	0	0	1	0	1	1	0	1	0	1	0	1	0	0	0	1	0	0	1	1	1
13	1	0	1	0	0	1	1	0	1	0	1	0	0	1	1	1	0	0	0	1	1	0	1	1
14	0	1	1	1	0	1	0	1	1	1	1	0	1	0	1	1	0	0	1	0	1	0	1	1
15	1	0	1	1	0	1	1	0	1	1	1	0	1	1	0	0	0	0	0	1	1	1	1	1
16	1	1	0	0	0	1	1	1	0	0	1	0	1	0	1	1	1	1						
17	1	1	0	0	1	1	1	1	0	1	0	0												
18	1	1	0	1	0	1	1	1	0	1	1	0												
19	1	1	1	0	1	1	1	1	0	1	1	1												
20	1	1	1	0	0	1	1	1	1	0	1	0												
21	1	1	1	1	0	0	1	1	1	0	1	1												
22	1	1	1	1	0	1	1	1	1	1	1	0												
23	1	1	1	1	1	1																		

xxxx01 and xxxx10 codify the same integer number

16 pairs of strings xxxx01 – xxxx10

Symmetry of xxxx01 ←→ xxxx10 exchange

Represented number	Length 6 binary strings																							
	8	7	4	2	1	1	8	7	4	2	1	1	8	7	4	2	1	1	8	7	4	2	1	1
0	0	0	0	0	0	0																		
1	0	0	0	0	0	1	0	0	0	0	1	0												
2	0	0	0	0	1	1	0	0	0	1	0	0												
3	0	0	0	1	0	1	0	0	0	1	1	0												
4	0	0	1	0	0	0	0	0	0	1	1	1												
5	0	0	1	0	0	1	0	0	1	0	1	0												
6	0	0	1	1	0	0	0	0	1	0	1	1												
7	0	0	1	1	0	1	0	0	1	1	1	0	0	1	0	0	0	0						
8	0	1	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	1	1	1
9	1	0	0	0	0	1	1	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0	1	1
10	0	1	0	1	0	1	0	1	0	1	1	0	1	0	0	1	0	0	1	0	0	0	1	1
11	1	0	0	1	0	1	1	0	0	1	1	0	0	1	1	0	0	0	0	1	0	1	1	1
12	0	1	1	0	0	1	0	1	1	0	1	0	1	0	1	0	0	0	1	0	0	1	1	1
13	1	0	1	0	0	1	1	0	1	0	1	0	0	1	1	1	0	0	0	1	1	0	1	1
14	0	1	1	1	0	1	0	1	1	1	1	0	1	0	1	1	0	0	1	0	1	0	1	1
15	1	0	1	1	0	1	1	0	1	1	1	0	1	1	0	0	0	0	0	1	1	1	1	1
16	1	1	0	0	0	1	1	1	0	0	1	0	1	0	1	1	1	1						
17	1	1	0	0	1	1	1	1	0	1	0	0												
18	1	1	0	1	0	1	1	1	0	1	1	0												
19	1	1	1	0	1	1	1	1	0	1	1	1												
20	1	1	1	0	0	1	1	1	1	0	1	0												
21	1	1	1	1	0	0	1	1	1	0	1	1												
22	1	1	1	1	0	1	1	1	1	1	1	0												
23	1	1	1	1	1	1																		

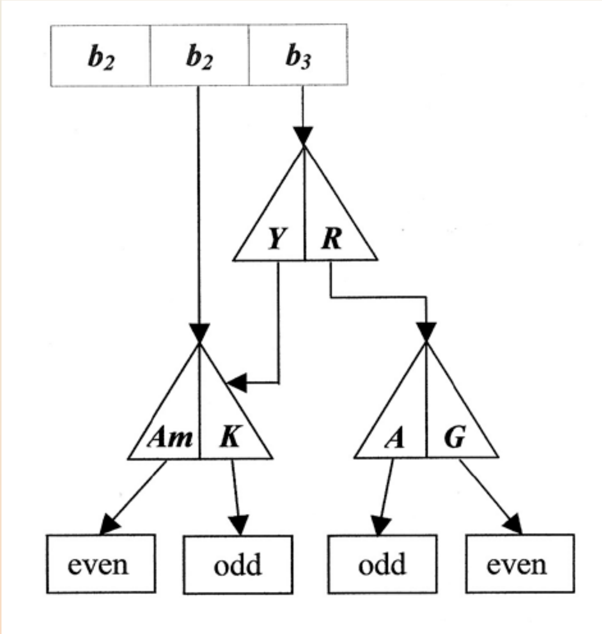
xxxx01 and xxxx10 codify the same integer number

16 pairs of strings xxxx01 – xxxx10

Parity Coding

The parity of the strings can be described in terms of biochemical properties

		U			C			A			G		
U	1	000001	Phe	15	101101	Ser	18	110110	Tyr	16	110010	Cys	U
	1	000010	Phe	15	101110	Ser	18	110101	Tyr	16	110001	Cys	C
	4	001000	Leu	15	011111	Ser	2	000100	Ter	16	101111	Cys	A
	11	011000	Leu	15	110000	Ser	2	000011	Ter	23	111111	Trp	G
C	11	100101	Leu	14	011110	Pro	3	000101	His	12	011010	Arg	U
	11	100110	Leu	14	011101	Pro	3	000110	His	12	011001	Arg	C
	4	000111	Leu	14	101100	Pro	17	110100	Gln	19	111000	Arg	A
	11	010111	Leu	14	101011	Pro	17	110011	Gln	12	101000	Arg	G
A	7	001101	Ile	8	010010	Thr	5	001001	Asn	22	111110	Ser	U
	7	001110	Ile	8	010001	Thr	5	001010	Asn	22	111101	Ser	C
	7	010000	Ile	8	100000	Thr	21	111011	Lys	19	110111	Arg	A
	0	000000	Met	8	001111	Thr	21	111100	Lys	12	100111	Arg	G
G	13	101001	Val	9	100001	Ala	20	111010	Asp	10	010110	Gly	U
	13	101010	Val	9	100010	Ala	20	111001	Asp	10	010101	Gly	C
	13	011100	Val	9	010011	Ala	6	001011	Glu	10	100011	Gly	A
	13	011011	Val	9	010100	Ala	6	001100	Glu	10	100100	Gly	G



Parity of a string = parity of the # of 1's, i.e., 0,1,1,0,1,0 = odd string

The complete set of parity rules:

Codons ending in A are odd (grey boxes); codons ending in G are even (white boxes)

Codons ending in T or C are odd if the second letter is T or G (Keto)

Codons ending in T or C are even if the second letter is A or C (Amino)

Dichotomic classes: Parity, Rumer, Hidden

Actual sequences can be binary coded and real sequences explored by **statistical methods**:

short-range correlations, **circular cod**es, etc.

Vertebrate mitochondrial genetic code

	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Trp	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Met	ACA Thr	AAA Lys	AGA Stop	A
	AUG Met	ACG Thr	AAG Lys	AGG Stop	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

Distribution of Degeneracy Vertebrate Mitochondrial	
Degeneracy number of synonymous codons	**Number of amino acids sharing this degeneracy**
2	**16**
4	**8**

Is it possible to describe the degeneracy distribution of the vertebrate mitochondrial genetic code by means of a non-power number representation?

In yellow are evidenced the differences with the standard nuclear genetic code

The 8,8,4,2,1,0 non-power solution for mitochondrial degeneracy

Represented number	Length 6 binary strings																							
	8	8	4	2	1	0	8	8	4	2	1	0	8	8	4	2	1	0	8	8	4	2	1	0
0	0	0	0	0	0	0							0	0	0	0	0	1						
1	0	0	0	0	1	0							0	0	0	0	1	1						
2	0	0	0	1	0	0							0	0	0	1	0	1						
3	0	0	0	1	1	0							0	0	0	1	1	1						
4	0	0	1	0	0	0							0	0	1	0	0	1						
5	0	0	1	0	1	0							0	0	1	0	1	1						
6	0	0	1	1	0	0							0	0	1	1	0	1						
7	0	0	1	1	1	0							0	0	1	1	1	1						
8	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1
9	1	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	1	1	0	1	0	0	1	1
10	1	0	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	1	0	1	0	1	0	1
11	1	0	0	1	1	0	0	1	0	1	1	0	1	0	0	1	1	1	0	1	0	1	1	1
12	1	0	1	0	0	0	0	1	1	0	0	0	1	0	1	0	0	1	0	1	1	0	0	1
13	1	0	1	0	1	0	0	1	1	0	1	0	1	0	1	0	1	1	0	1	1	0	1	1
14	1	0	1	1	0	0	0	1	1	1	0	0	1	0	1	1	0	1	0	1	1	1	0	1
15							0	1	1	1	1	0	1	0	1	1	1	1	0	1	1	1	1	1
16							1	1	0	0	0	0							1	1	0	0	0	1
17							1	1	0	0	1	0							1	1	0	0	1	1
18							1	1	0	1	0	0							1	1	0	1	0	1
19							1	1	0	1	1	0							1	1	0	1	1	1
20							1	1	1	0	0	0							1	1	1	0	0	1
21							1	1	1	0	1	0							1	1	1	0	1	1
22							1	1	1	1	0	0							1	1	1	1	0	1
23							1	1	1	1	1	0							1	1	1	1	1	1

Degeneracy	Number
Non-Power representation	
2	16
4	8
Mitochondrial Genetic Code	
2	16
4	8
Euplotes Nuclear Genetic Code	
1	2
2	12
3	2
4	8

Symmetry of A ←→ G exchange (Mitochondrial Code)

NNA and NNG codify the same amino acid

16 pairs of codons NNA - NNG

	U		C		A		G		
U	Phe	UUU	Ser	UCU	Tyr	UAU	Cys	UGU	U
		UUC		UCC		UAC		UGC	C
	Leu	UUA		UCA	Stop	UAA	Trp	UGA	A
		UUG		UCG		UAG		UGG	G
C	Leu	CUU	Pro	CCU	His	CAU	Arg	CGU	U
		CUC		CCC		CAC		CGC	C
		CUA		CCA	Gln	CAA		CGA	A
		CUG		CCG		CAG		CGG	G
A	Ile	AUU	Thr	ACU	Asn	AAU	Ser	AGU	U
		AUC		ACC		AAC		AGC	C
	Met	AUA		ACA	Lys	AAA	Stop	AGA	A
		AUG		ACG		AAG		AGG	G
G	Val	GUU	Ala	GCU	Asp	GAU	Gly	GGU	U
		GUC		GCC		GAC		GGC	C
		GUA		GCA	Glu	GAA		GGA	A
		GUG		GCG		GAG		GGG	G

Common to many mitochondrial versions of the genetic code

Symmetry of A ←→ G exchange (Mitochondrial Code)

NNA and NNG codify the same amino acid

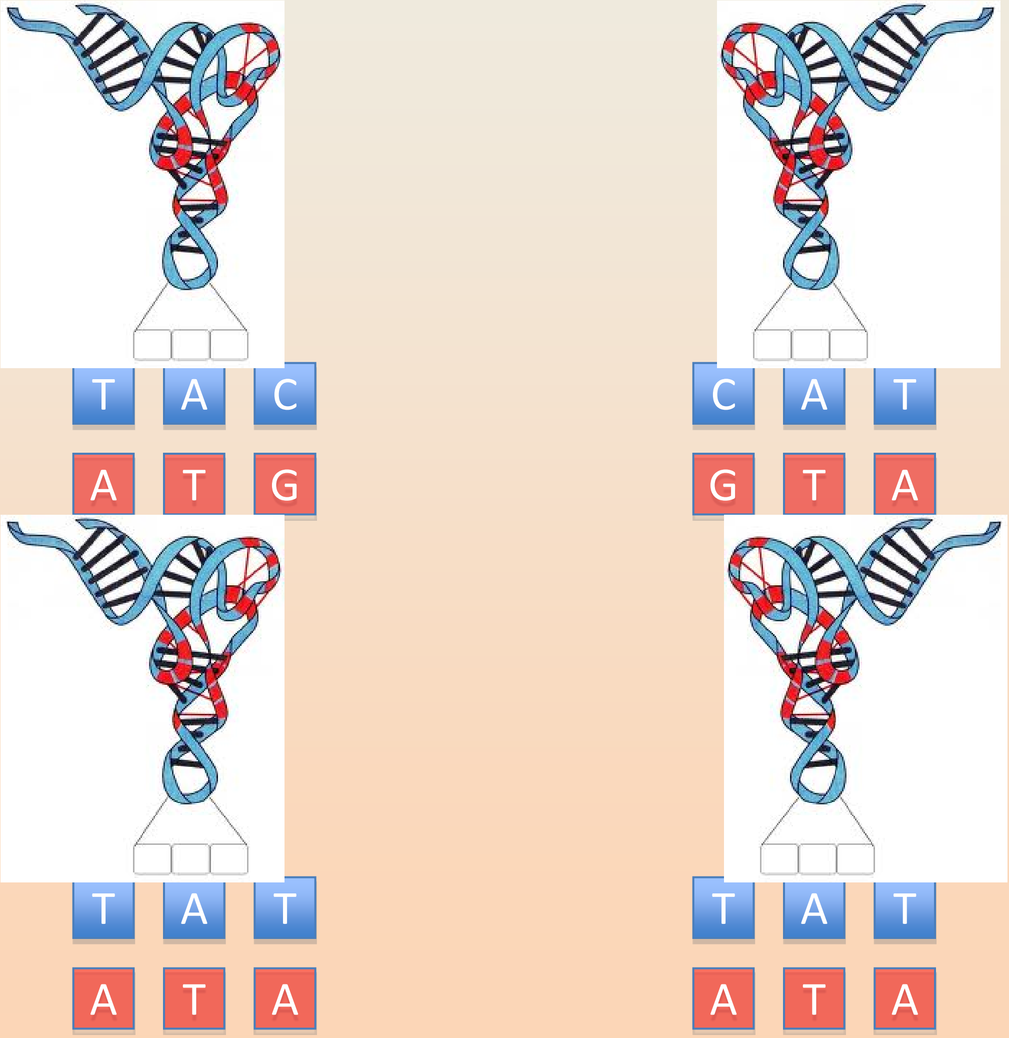
16 pairs of codons NNA - NNG

	U		C		A		G		
U	Phe	UUU	Ser	UCU	Tyr	UAU	Cys	UGU	U
		UUC		UCC		UAC		UGC	C
	Leu	UUA		UCA	Stop	UAA	Trp	UGA	A
		UUG		UCG		UAG		UGG	G
C	Leu	CUU	Pro	CCU	His	CAU	Arg	CGU	U
		CUC		CCC		CAC		CGC	C
		CUA		CCA	Gln	CAA		CGA	A
		CUG		CCG		CAG		CGG	G
A	Ile	AUU	Thr	ACU	Asn	AAU	Ser	AGU	U
		AUC		ACC		AAC		AGC	C
	Met	AUA		ACA	Lys	AAA	Stop	AGA	A
		AUG		ACG		AAG		AGG	G
G	Val	GUU	Ala	GCU	Asp	GAU	Gly	GGU	U
		GUC		GCC		GAC		GGC	C
		GUA		GCA	Glu	GAA		GGA	A
		GUG		GCG		GAG		GGG	G

Common to many mitochondrial versions of the genetic code

Origin of degeneracy in amino acid coding

Hypothesis of reversible ancient tRNAs



S. Nikolajewa. M. Friedel, A. Beyer, and T. Wilhelm, The new classification scheme of the genetic code, its early evolution, and tRNA usage, Journal of Bioinformatics and Computational Biology, 12, 54 (2005) 1-12

Half the Degeneracy of the Mitochondrial Genetic Code

Represented number												
		8	8	4	2	1		8	8	4	2	1
0		0	0	0	0	0						
1		0	0	0	0	1						
2		0	0	0	1	0						
3		0	0	0	1	1						
4		0	0	1	0	0						
5		0	0	1	0	1						
6		0	0	1	1	0						
7		0	0	1	1	1						
8		1	0	0	0	0		0	1	0	0	0
9		1	0	0	0	1		0	1	0	0	1
10		1	0	0	1	0		0	1	0	1	0
11		1	0	0	1	1		0	1	0	1	1
12		1	0	1	0	0		0	1	1	0	0
13		1	0	1	0	1		0	1	1	0	1
14		1	0	1	1	0		0	1	1	1	0
15		1	0	1	1	1		0	1	1	1	1
16								1	1	0	0	0
17								1	1	0	0	1
18								1	1	0	1	0
19								1	1	0	1	1
20								1	1	1	0	0
21								1	1	1	0	1
22								1	1	1	1	0
23								1	1	1	1	1

Half the degeneracy of the genetic code Obtained by excluding the 0 non-power base of the representation

(8,8,4,2,1)

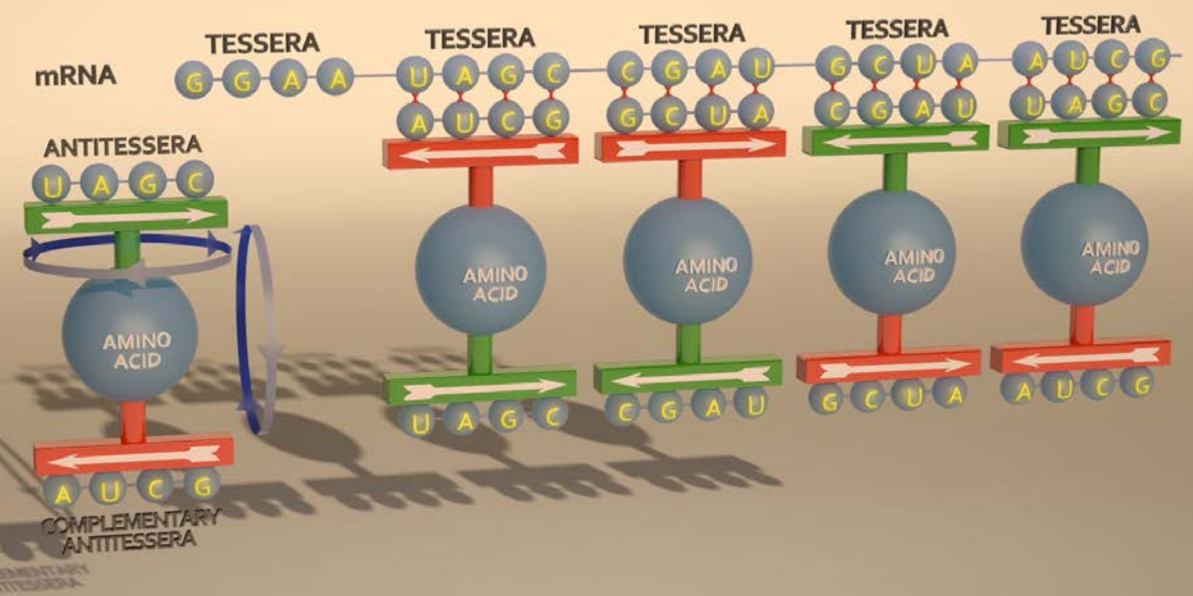
The Complete Tesserae Representation

	A	A	A	A	I	A	A	U	U	C	A	A	G	G	Y/R	A	A	C	C	R
	U	U	U	U		U	U	A	A		U	U	C	C		U	U	G	G	
	C	C	C	C		C	C	G	G		C	C	U	U		C	C	A	A	
	G	G	G	G		G	G	C	C		G	G	A	A		G	G	U	U	
	A	U	U	A	C	A	U	A	U	I	A	U	C	G	R	A	U	G	C	Y/R
	U	A	A	U		U	A	U	A		U	A	G	C		U	A	C	G	
	C	G	G	C		C	G	C	G		C	G	A	U		C	G	U	A	
	G	C	C	G		G	C	G	C		G	C	U	A		G	C	A	U	
	A	G	G	A	Y/R	A	G	C	U	R	A	G	A	G	I	A	G	U	C	C
	U	C	C	U		U	C	G	A		U	C	U	C		U	C	A	G	
	C	U	U	C		C	U	A	G		C	U	C	U		C	U	G	A	
	G	A	A	G		G	A	U	C		G	A	G	A		G	A	C	U	
	A	C	C	A	R	A	C	G	U	Y/R	A	C	U	G	C	A	C	A	C	I
	U	G	G	U		U	G	C	A		U	G	A	C		U	G	U	G	
	C	A	A	C		C	A	U	G		C	A	G	U		C	A	C	A	
	G	U	U	G		G	U	A	C		G	U	C	A		G	U	G	U	

Group properties of the four symmetric transformations, i.e., I, C, YR and KM

Klein 4-group V

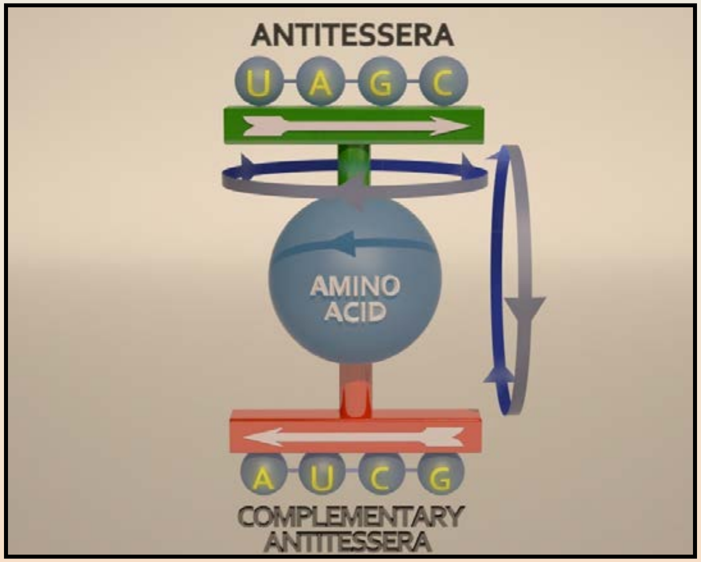
Origin of degeneracy in amino acid coding



Degeneracy as plurality of states characterized by the same interaction energy

D.L. Gonzalez, S. Giannerini, and R. Rosa, On the origin of the mitochondrial genetic code: Towards a Unified mathematical framework for the management of genetic information, Nature Precedings, <http://dx.doi.org/10.1038/npre.2012.7136.1> (2012)
H. Seligman, Pocketknife tRNA hypothesis: anticodons in mammal mitochondrial tRNA side-arm loops translate proteins? Biosystems (2013)

Origin of the mitochondrial genetic code



	A	A	A	A	I	A	A	U	U	C	A	A	G	G	Y/R	A	A	C	C	R
	U	U	U	U		U	U	A	A		U	U	C	C		U	U	G	G	
	C	C	C	C		C	C	G	G		C	C	U	U		C	C	A	A	
	G	G	G	G		G	G	C	C		G	G	A	A		G	G	U	U	
	A	U	U	A	C	A	U	A	U	I	A	U	C	G	R	A	U	G	C	Y/R
	U	A	A	U		U	A	U	A		U	A	G	C		U	A	C	G	
	C	G	G	C		C	G	C	G		C	G	A	U		C	G	U	A	
	G	C	C	G		G	C	G	C		G	C	U	A		G	C	A	U	
	A	G	G	A	Y/R	A	G	C	U	R	A	G	A	G	I	A	G	U	C	C
	U	C	C	U		U	C	G	A		U	C	U	C		U	C	A	G	
	C	U	U	C		C	U	A	G		C	U	C	U		C	U	G	A	
	G	A	A	G		G	A	U	C		G	A	G	A		G	A	C	U	
	A	C	C	A	R	A	C	G	U	Y/R	A	C	U	G	C	A	C	A	C	I
	U	G	G	U		U	G	C	A		U	G	A	C		U	G	U	G	
	C	A	A	C		C	A	U	G		C	A	G	U		C	A	C	A	
	G	U	U	G		G	U	A	C		G	U	C	A		G	U	G	U	

Distribution of Degeneracy Vertebrate Mitochondrial	
Degeneracy number of synonymous codons	**Number of amino acids sharing this degeneracy**
2	**16**
4	**8**

Distribution of Degeneracy Non-power system	
8,8,4,2,1,0	
Degeneracy	**# of integers**
2	**16**
4	**8**

Distribution of Degeneracy Tesserae	
Degeneracy Number of synonymous tesserae	**Number of amino acids sharing this degeneracy**
2	**16**
4	**8**

Tesserae Properties

	A	A	A	A	I	A	A	U	U	C	A	A	G	G	Y/R	A	A	C	C	R
	U	U	U	U		U	U	A	A		U	U	C	C		U	U	G	G	
	C	C	C	C		C	C	G	G		C	C	U	U		C	C	A	A	
	G	G	G	G		G	G	C	C		G	G	A	A		G	G	U	U	
	A	U	U	A	C	A	U	A	U	I	A	U	C	G	R	A	U	G	C	Y/R
	U	A	A	U		U	A	U	A		U	A	G	C		U	A	C	G	
	C	G	G	C		C	G	C	G		C	G	A	U		C	G	U	A	
	G	C	C	G		G	C	G	C		G	C	U	A		G	C	A	U	
	A	G	G	A	Y/R	A	G	C	U	R	A	G	A	G	I	A	G	U	C	C
	U	C	C	U		U	C	G	A		U	C	U	C		U	C	A	G	
	C	U	U	C		C	U	A	G		C	U	C	U		C	U	G	A	
	G	A	A	G		G	A	U	C		G	A	G	A		G	A	C	U	
	A	C	C	A	R	A	C	G	U	Y/R	A	C	U	G	C	A	C	A	C	I
	U	G	G	U		U	G	C	A		U	G	A	C		U	G	U	G	
	C	A	A	C		C	A	U	G		C	A	G	U		C	A	C	A	
	G	U	U	G		G	U	A	C		G	U	C	A		G	U	G	U	

A,A,A,A -> A,A,T,A

Point mutation immunity, +1 Frame-shift immunity, Rumer’s transformation

Conserved Features of the Mitochondrial Genetic Code

Genetic Code

	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Trp	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Met	ACA Thr	AAA Lys	AGA Stop	A
	AUG Met	ACG Thr	AAG Lys	AGG Stop	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

Non-power Representation

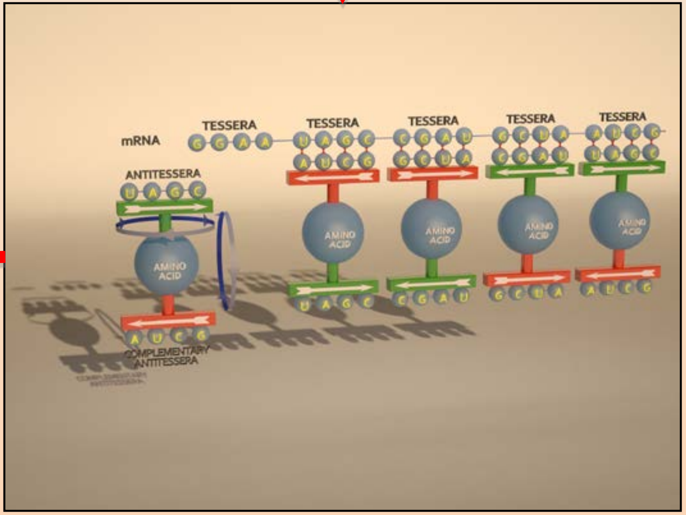
Represented number	Length 6 binary strings			
	8 8 4 2 1 0	8 8 4 2 1 0	8 8 4 2 1 0	8 8 4 2 1 0
0	0 0 0 0 0 0		0 0 0 0 0 1	
1	0 0 0 0 1 0		0 0 0 0 1 1	
2	0 0 0 1 0 0		0 0 0 1 0 1	
3	0 0 0 1 1 0		0 0 0 1 1 1	
4	0 0 1 0 0 0		0 0 1 0 0 1	
5	0 0 1 0 1 0		0 0 1 0 1 1	
6	0 0 1 1 0 0		0 0 1 1 0 1	
7	0 0 1 1 1 0		0 0 1 1 1 1	
8	1 0 0 0 0 0	0 1 0 0 0 0	1 0 0 0 0 1	0 1 0 0 0 1
9	1 0 0 0 1 0	0 1 0 0 1 0	1 0 0 0 1 1	0 1 0 0 1 1
10	1 0 0 1 0 0	0 1 0 1 0 0	1 0 0 1 0 1	0 1 0 1 0 1
11	1 0 0 1 1 0	0 1 0 1 1 0	1 0 0 1 1 1	0 1 0 1 1 1
12	1 0 1 0 0 0	0 1 1 0 0 0	1 0 1 0 0 1	0 1 1 0 0 1
13	1 0 1 0 1 0	0 1 1 0 1 0	1 0 1 0 1 1	0 1 1 0 1 1
14	1 0 1 1 0 0	0 1 1 1 0 0	1 0 1 1 0 1	0 1 1 1 0 1
15		0 1 1 1 1 0	1 0 1 1 1 1	0 1 1 1 1 1
16		1 1 0 0 0 0		1 1 0 0 0 1
17		1 1 0 0 1 0		1 1 0 0 1 1
18		1 1 0 1 0 0		1 1 0 1 0 1
19		1 1 0 1 1 0		1 1 0 1 1 1
20		1 1 1 0 0 0		1 1 1 0 0 1
21		1 1 1 0 1 0		1 1 1 0 1 1
22		1 1 1 1 0 0		1 1 1 1 0 1
23		1 1 1 1 1 0		1 1 1 1 1 1

Tesserae (di-dinucleotides)

	A	A	A	A		A	A	U	U		A	A	G	G		A	A	C	C	
	U	U	U	U	I	U	U	A	A	C	U	U	C	C	Y/R	U	U	G	G	R
	C	C	C	C		C	C	G	G		C	C	U	U		C	C	A	A	
	G	G	G	G		G	G	C	C		G	G	A	A		G	G	U	U	
	A	U	U	A		A	U	A	U		A	U	C	G		A	U	G	C	
	U	A	A	U	C	U	A	U	A	I	U	A	G	C	R	U	A	C	G	Y/R
	C	G	G	C		C	G	C	G		C	G	A	U		C	G	U	A	
	G	C	C	G		G	C	G	C		G	C	U	A		G	C	A	U	
	A	G	G	A		A	G	C	U		A	G	A	G		A	G	U	C	
	U	C	C	U	Y/R	U	C	G	A	R	U	C	U	C	I	U	C	A	G	C
	C	U	U	C		C	U	A	G		C	U	C	U		C	U	G	A	
	G	A	A	G		G	A	U	C		G	A	G	A		G	A	C	U	
	A	C	C	A		A	C	G	U		A	C	U	G		A	C	A	C	
	U	G	G	U	R	U	G	C	A	Y/R	U	G	A	C	C	U	G	U	G	I
	C	A	A	C		C	A	U	G		C	A	G	U		C	A	C	A	
	G	U	U	G		G	U	A	C		G	U	C	A		G	U	G	U	

Symmetries

Degeneracy Distribution



Putative Ancient tRNA Adaptors

Conserved Features

Why these?

Publications

"Una Nuova Descrizione Matematica del Codice Genetico" D.L. Gonzalez and M. Zanna, **Systema Naturae, Annali di Biologia Teorica, V. 5, pp. 219-236, (2003).**
"Can the genetic code be mathematically described?", Diego L. Gonzalez, **Medical Science Monitor, v.10, n.4, HY11-17 (2004).**
"Detecting Structure in Parity Binary Sequences: Error Correction and Detection in DNA", Diego Luis Gonzalez, Simone Giannerini and Rodolfo Rosa, **IEEE-EMB Magazine, Special Issue on Communication Theory, Coding Theory and Molecular Biology,** pp. 69-81, January/February (2006).
"Error Detection and Correction Codes" D.L. Gonzalez, in **"The Codes of Life", chapter 6**, Marcello Barbieri Editor, **Springer Verlag (2008).**
"The Mathematical Structure of the Genetic Code" D.L. Gonzalez, in **"The Codes of Life", chapter 17**, Marcello Barbieri Editor, **Springer Verlag (2008).**
"Strong Short-Range Correlations and Codon Class Dichotomies classes in Coding DNA Sequences", Diego L. Gonzalez, Simone Giannerini and Rodolfo Rosa, **Physical Review E,78, n.5, 08481, (2008).**
"The mathematical structure of the genetic code: a tool for inquiring on the origin of life" Diego L. Gonzalez, S. Giannerini, and R. Rosa, **Rivista di Statistica, Anno LXIX, n. 2-3, pp. 143-158,(2009).**
"Circular codes revisited: a statistical approach" Diego Luis Gonzalez, Simone Giannerini and Rodolfo Rosa, **Journal of Theoretical Biology, vol. 275, n.1, pp21-28, (2011).**
"DNA, circular codes and dichotomic classes: a quasi-crystal framework", Diego Luis Gonzalez, Simone Giannerini and Rodolfo Rosa, **Phil Trans A of the Royal Society,** Beyond Crystals: the dialectic of information and structure, theme issue, **370, 2987-3006; doi:10.1098/rsta.2011.0387, (2012).**
"On the origin of the mitochondrial genetic code: Towards a unified mathematical framework for the management of genetic information" Gonzalez, Diego Luis, Giannerini, Simone, and Rosa, Rodolfo. Available from Nature Precedings **<http://dx.doi.org/10.1038/npre.2012.7136.1> (2012)**
"Genome characterization through dichotomic classes: an analysis of the whole chromosome 1 of A. Thaliana, Enrico Properzi, Simone Giannerini, Diego L. Gonzalez, and Rodolfo Rosa, **Mathematical Biosciences and Engineering, v. 10, n. 1, pp 199-219, (2013).**