S. V. Kozyrev

Steklov Mathematical Institute

p–Adic numbers in genetics

The Third International Conference on p–Adic Mathematical Physics

October 1 – 6, 2007

A.Yu. Khrennikov, S.V. Kozyrev, Genetic code on the dyadic plane.
Physica A: Statistical Mechanics and its Applications.
2007. V.381. P.265-272.

http://arxiv.org/abs/q-bio.QM/0701007

The genetic code has an intrinsic p–adic structure

of 2–adic plane

(possesses 2–dimensional 2–adic norm)

The genetic code

DNA (RNA) — finite sequence (linear polymer) of nucleotides
protein — finite sequence of amino acids

Nucleotides: C, A, T (or U), G
Cytosine, Adenine, Thymine (or Uracil), Guanine
Triplet structure of genetic code
Codon — triple $C_1C_2C_3$ of nucleotides
Genetic code — maps codons onto amino acids

Encode 20 amino acids and 1 stop symbol
4 nucleotides, $4 \times 4 \times 4 = 64$ codons
$64 > 21$ — problem of degeneracy of genetic code

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

Table 1 : the vertebrate mitochondrial genetic code

The first application of ultrametric methods

Carl von Linne (Växjö, Sweden), Systema Naturae, 1735

Genetic code and coding theory

Gray codes and error correcting codes in relation to genetic code:

R.Swanson, A unifying concept for the amino acid code,
Bulletin of Mathematical Biology, 1984. V.46. N.2. P.187-203.

Error correction by redundancy of the code
degeneracy of the genetic code provides partial error correction

p–Adic methods in genetics

Description of the genetic code
by Cantor–like subset of 5–adic ultrametric space
with two metrics, 5–adic and 2–adic
(B.Dragovich, A.Dragovich, 2006)

2–dimensional 2–adic parametrization of genetic code
$64 = 8 \times 8$ — dyadic plane
genetic chessboard
(A.Yu.Khrennikov, S.V.Kozyrev, 2007)

Dyadic plane of codons

1) Parameterize the nucleotides by pairs of 0,1

$$\begin{array}{|c|c|}\hline A & G \\ \hline U & C \\ \hline \end{array} = \begin{array}{|c|c|}\hline 00 & 01 \\ \hline 10 & 11 \\ \hline \end{array} \qquad (A)$$

Chemical meaning:
1-st line — purines
2-nd line — pyrimidines
1-st column — weak H-bond
2-nd column — strong H-bond

2) The order of nucleotides in the codon

$$2 > 1 > 3 \qquad (B)$$

3) Dyadic plane — square 8×8 with the coordinates (x, y):

$$x = (x_0 x_1 x_2) = x_0 + 2x_1 + 4x_2, \quad x_i = 0, 1,$$
$$y = (y_0 y_1 y_2) = y_0 + 2y_1 + 4y_2, \quad y_i = 0, 1.$$

ρ maps the codon into the point of the dyadic plane

$$\rho : C_1 C_2 C_3 \mapsto (x, y) = (x_0 x_1 x_2, y_0 y_1 y_2),$$

C_2 defines the pair (x_0, y_0),
C_1 defines the pair (x_1, y_1),
C_3 defines the pair (x_2, y_2).

Nucleotides define the pairs of digits according to the rule (A), the scale (2–adic norm) of the pair is given by the rule (B).

Enumerate the lines and the columns of the dyadic plane:

$$\eta : x \mapsto \widetilde{x}, \qquad y \mapsto \widetilde{y};$$
$$\eta : x_0 + 2x_1 + 4x_2 \mapsto 1 + 4x_0 + 2x_1 + x_2;$$
$$\eta : y_0 + 2y_1 + 4y_2 \mapsto 1 + 4y_0 + 2y_1 + y_2.$$

Equivalently:

$$\eta : 0, 4, 2, 6, 1, 5, 3, 7 \mapsto 1, 2, 3, 4, 5, 6, 7, 8.$$

Dyadic plane of codons

AAA	AAG	GAA	GAG	AGA	AGG	GGA	GGG
AAU	AAC	GAU	GAC	AGU	AGC	GGU	GGC
UAA	UAG	CAA	CAG	UGA	UGG	CGA	CGG
UAU	UAC	CAU	CAC	UGU	UGC	CGU	CGC
AUA	AUG	GUA	GUG	ACA	ACG	GCA	GCG
AUU	AUC	GUU	GUC	ACU	ACC	GCU	GCC
UUA	UUG	CUA	CUG	UCA	UCG	CCA	CCG
UUU	UUC	CUU	CUC	UCU	UCC	CCU	CCC

with 2–dimensional 2–adic metric

$$d(C_1C_2C_3, C_1'C_2'C_3') = \max(|x - x'|_2, |y - y'|_2)$$

$$(x, y) = \rho(C_1C_2C_3), \quad (x', y') = \rho(C_1'C_2'C_3')$$

the distance can take values 1, 1/2, 1/4.

Apply to the dyadic plane of codons the Vertebrate Mitochondrial Code:

Lys / Asn	Glu / Asp	Ter / Ser	Gly
Ter / Tyr	Gln / His	Trp / Cys	Arg
Met / Ile	Val	Thr	Ala
Leu / Phe	Leu	Ser	Pro

In particular

$$\begin{array}{|c|c|}\hline AAA & AAG \\ \hline AAU & AAC \\ \hline \end{array} \rightarrow \boxed{\begin{array}{c}\text{Lys} \\ \text{Asn}\end{array}}, \qquad \begin{array}{|c|c|}\hline CCA & CCG \\ \hline CCU & CCC \\ \hline \end{array} \rightarrow \boxed{\text{Pro}}$$

Very regular distribution of degeneracy

The map of the genetic code:

1) 1/4 – Locally constant w.r.t. the horizontal coordinate;

2) 1/4 – Locally constant w.r.t. the horizontal
and vertical coordinates on the half of the dyadic plane;

vertical coordinate (i.e. the difference between purines and pyrimidines) is more important

3a) Degeneracies in 1/4 – subballs
of upper left 1/2 – ball are identical;

3b) Degeneracies in 1/4 – subballs
of lower right 1/2 – ball are identical;

3c) Degeneracies in upper right and lower left
1/2 – balls are identical
(the dyadic plane is symmetrical);

physical–chemical properties (hydrophobicity, polarity)
are clusterized in the dyadic plane

The table of hydrophobic amino acids

—	—	—	
—	—	Trp / Cys	
Met / Ile	Val		
Leu / Phe	Leu		

hydrophobic amino acids are clusterized in two balls

similar amino acids are 2–adically close in the dyadic plane

Comparing different versions of genetic code
we make conjectures about evolution (more regular code is older)
The eucaryotic code is less regular

Lys / Asn	Glu / Asp	Arg / Ser	Gly
Ter / Tyr	Gln / His	Ter|Trp / Cys	Arg
Ile|Met / Ile	Val	Thr	Ala
Leu / Phe	Leu	Ser	Pro

Theorem
degeneracy of the genetic code — 8×8 **2–adic local constancy**

1) The genetic code is a locally constant map on the dyadic plane;

2) The degeneracy of the genetic code is regularly distributed in the dyadic plane;

3) Amino acids with similar physical–chemical properties are 2–adically close in the dyadic plane.