Seventh International Conference in Code Biology

Lužnica (Croatia) 31 August – 4 September 2021

ABSTRACTS

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

Metabolic closure of the system of nucleotides and the origin of biological codes

Abir U. Igamberdiev
Department of Biology, Memorial University of Newfoundland
St. John's, NL, A1B 3X9, Canada
igamberdiev@mun.ca

The general structure of metabolism includes the reproduction of catalysts that govern metabolism. In this structure the system becomes closed to efficient causation as defined by Robert Rosen. The maintenance and operation of the catalysts takes place via the set of free nucleotides while the synthesis of catalysts occurs via the information encoded by the set of nucleotides arranged in polymers of RNA and DNA. Both energy charge and genetic information use the components of the same pool of nucleoside triphosphates which is equilibrated by thermodynamic buffering enzymes such as nucleoside diphosphate kinase, adenylate kinase, etc. This occurs in a way that the system becomes internally stable and metabolically closed. Metabolic closure is a prerequisite for the internal autonomy of biological systems. The function of ATP, GTP, UTP, and CTP is dual, as these species participate both in the general metabolism as free nucleotides and in the transfer of genetic information via covalent polymerization to nucleic acids. The changes in their pools directly impact both bioenergetic pathways and nucleic acid turnover. The main challenge is a difficulty of formal joint representation of energetic and information fluxes in the frames of a whole system in biological theories. The formal unified representation of energetic and information fluxes that involve the same pool of components is an important task for computational biology. The operation of the genetic and the metabolic components of the system generate simple computable rules within the systems that are established via the complex network of equilibrium and non-equilibrium enzymatic reactions. Biological systems are organized in such a way that deviations from the equilibrium tend to return to the initial stable non-equilibrium state that characterizes metabolic structure of biological systems. Operation of biological systems as autonomous and self-regulated is achieved via local equilibria in the closed cyclic structure of metabolism where the same nucleotide types act as coenzymes and as the components of nucleic acids. The systems of feedbacks from nucleoside phosphate equilibria make it possible for metabolism itself and for the process of replacement of its catalysts and matrices to operate as a whole entity, with the established computational principles. Biological codes appear as the consequence of metabolic closure. They provide a possibility of self-maintenance and reproduction of biological system via the arrangement of fluxes of energy and information spatially and temporally and via the introduction of internal computational principles within biological systems.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

The Major Evolutionary Transitions and Codes of Life

Ádám Kun
Theoretical Biology and Evolutionary Ecology Research Group, Budapest, Hungary
Institute for Advanced Studies Kőszeg, Kőszeg, Hungary
Parmenides Centre for the Conceptual Foundation of Science, Munich, Germany
kunadam@elte.hu

The major evolutionary transitions as well as the evolution of the codes of life are key elements in macroevolution which are characterized by the increase in complexity Major evolutionary transitions ensues by a transition in individuality and by the evolution of a novel mode of using, transmitting or storing information. Here is where codes of life enter the picture: they are arbitrary mappings between different (mostly) molecular species. This flexibility allows information to be employed in a variety of ways, which can fuel evolutionary innovation. The collation of the list of major evolutionary transitions and the list of codes of life show a clear pattern: codes evolved prior to a major evolutionary transition and then played roles in the transition and/or in transformation of the new individual. The evolution of codes of life is in themselves not major evolutionary transitions but allow major evolutionary transitions to happen. This could help us to identify new organic codes.

Seventh International Conference in Code Biology - Lužnica (Croatia) 31 August – 4 September 2021

The complexity, and hierarchy of languages for the hereditary information unfolding in development

Alexander V. Spirov
Institute of Scientific Information for Social Sciences
Russian Academy of Sciences, Moscow, Russia
alexander.spirov@gmail.com

In this communication, we consider the principles of the genetic information unfolding in a multicellular embryo. We focus on informational aspects of the gene activity regulation. Genetic information unfolds even before fertilization and even after death. The regulation of genes is carried out by codes (or “languages”) of regulation. At least four languages for gene regulation can be distinguished. These are gene-gene interaction language (gene regulatory networks; targets are genes) (1); Cell-cell signaling language (final target is genes) (2); Language of long-distance biochemical regulators: growth factor signaling, hormones, etc. (final target is genes) (3); Languages of immune and nervous regulation: receptors, neurotransmitters (the final target is genes) (4).

The gene is not only the “carrier of genetic information,” but it typically includes an extensive (scattered along the chromosome) set of regulatory elements involved in the control of gene activity. At the basic level, genes are turned on or off when some proteins specifically bind to their DNA targets (direct regulations). In addition, other molecules are involved in gene regulation indirectly. The changing set of indirect regulatory inputs is understood as regulatory context. As a result of all regulatory signals, direct and indirect, the gene is activated at this time in a given cell.

It is crucial that the processes of gene regulation unfold not only in time, but also in space of the embryo. With an increase in the number of cells in the embryo, another level of control of gene activity appears. This is a cellular signaling: embryonic cells begin to exchange specific chemical "signals", which ultimately also leads to turning on or off genes in target cells. Accordingly, the context that determines the response of a gene to regulatory molecules extends beyond a single cell. A gene receives signals from its own cell, from neighboring cells, from distant cells, from the external environment. The gene processes continuous flows of information and makes decisions.

During embryogenesis, information is transmitted from the current biochemical prepattern to the future structures of the embryo through morphogenetic movements. However, the understanding of the process of the unfolding of genetic information in an individual development leads to conclusions about how complex and not obvious everything is.

In parallel with the multiplication of cells, epigenetic regulation mechanisms are unfolding. This is when nuclei, often neighboring cells, became able to modify differently the molecular machinery of chromatin (chemical modifications, first of all). This gives another level in the unfolding of genetic information, qualitatively complicating the spatial dimension of the intercellular connections of gene regulation.

Our conclusions: (1) In any organism, more than one language (code) is usually involved in the regulation of the process of unfolding hereditary information; these languages usually form a more complex system, changing in space and in time of the organism. (2) The systems of languages are typically hierarchical; We can distinguish a lower-level language, as well as higher-level languages. (3) The systems of languages and their hierarchy change not only in time of the organism, but also in the organismic space; each small part of a multicellular developing organism has its own version of the hierarchy of languages of regulation. We believe that the growing spatial distribution of the gene regulatory processes, the growing contextual dependence and the growing multiplicity and hierarchy of the regulatory “languages” require further in-depth study.

Seventh International Conference in Code Biology - Lužnica (Croatia) 31 August – 4 September 2021

Embodied Dialogue
Words as Deeds

Anna Aragno
National Association for the Advancement of Psychoanalysis
140 West End Avenue, New York, NY, 10023, USA
annaragno@earthlink.net

> *The life of mind is a totality of levels, which on one hand exist side by side, but which on the other, appear transitorily on after the other. The moments which the mind seems to have left behind actually exist in it at the present time in full depth.*
>
> Hegel "*Lectures on Philosophy of History*" (1949)
>
> Sämmtliche Werke 9

This presentation addresses the interpretive implications of referencing multi-code meanings in the framework of a revised multi-code, bio-semiotic, developmental model of mind and communication (Aragno, 1997/2016, 2008/2016). Freud tailored an interpretive dialogue, psycho-analysis, which he called his 'scientific method,' in order to gain maximum access in uncovering and interpreting unconscious processes and meanings. As a methodology designed specifically for this purpose its tilted, unconventional, interlocutory protocol and interpretive agendas, establish a complex multileveled, bio-semiotic semantic field triggering many regressive interactive processes that reveal pre- and non-linguistic communicative-forms. The specialized listening/observational stance required of this methodology also exposes embodied unconscious expressions that have little to do with surface linguistic content, to which it relates only metaphorically. In this bi-directional, dialogic orchestration, of multi-coded (body/mind) condensed complexity, wherein the sole interpretive instrument allowed is language, word-meanings may partially dissolve their semiotic structure expressing earlier, emotive-needs and underlying drive impulsions, thereby becoming as powerful as deeds, while exposing their antecedent embodied origins.

The renowned tree of life image is a metaphor, model, and research tool, used to explore the evolution of life and describe relationships between living organisms, as in a famous passage of Charles Darwin's 'On the Origin of Species' (1895). It's universal appeal, however, may be generalized to serve other epigenetic, multi-level and multi-coded conceptualizations and situations. l will make ample use of this natural iconic template to illustrate major underlying themes of my talk.

Populational approach to eukaryogenesis

Branimir Ećimović
Department of Biology, Faculty of Science and Education
University of Mostar, Bosnia and Herzegovina
branimir.ecimovic@gmail.com

Eukaryogenesis still stands as an open problem of macroevolutionary transition even though multiple well-respected theories exist. Despite the popularity and relative plausibility of Margulis's Serial endosymbiotic theory, this opinion is reductionist at best, and to holistically solve the problem different concept(s) should be applied. In a world dominated by prokaryotes, novel biological forms had to substantially increase their genetic and metabolic complexity in order to survive and thrive (Barbieri, 1985), and the appointment of multicellular LECA instead of single-celled progenitor at the basis of eukaryogenesis could present the conceptual solution (O'Malley et al. 2019). This way, it would be possible to conceptualize the parallel development of complex eukaryotic characteristics such as cell to cell communication, division of labor, metabolic and genomic novelties, etc. It has been recently proved that developmental process of bacterial biofilms is comparable to organismal development in higher eukaryotes (Futo et al. 2020), allowing us to assume that prokaryotic and (later) eukaryotic unicellular life evolved only secondary as a mechanism of dispersion. Signal transduction codes that governed the formation of ancient biofilms could mark the establishment of the coding process which later diverged into modern prokaryotic quorum sensing and eukaryotic embryotic development processes. Safer communal environment propagates extensive HGT and even genomic fusions, two processes in line with novel ring of life hypothesis (Riviera and Lake, 2004). Populational approach to eukaryogenesis could stand as a new paradigm for theoretical and empirical explorations, such as the correlations between the coding patterns of (archaeal-eubacterial) biofilms and multicellular eukaryotes.

References:

Barbieri, M. (2021). The semantic theory of evolution. Routledge.
Futo, M., Opašić, L., Koska, S., et al. (2021). Embryo-like features in developing Bacillus subtilis biofilms. Molecular biology and evolution, 38(1), 31-47.
O'Malley, M. A., Leger, M. M., Wideman, J. G., & Ruiz-Trillo, I. (2019). Concepts of the last eukaryotic common ancestor. Nature ecology & evolution, 3(3), 338-344.
Rivera, M. C., & Lake, J. A. (2004). The ring of life provides evidence for a genome fusion origin of eukaryotes. Nature, 431(7005), 152-155.

Seventh International Conference in Code Biology - Lužnica (Croatia) 31 August - 4 September 2021

Coding and Re-coding Emotions

Candice L. Shelby
Professor of Philosophy, University of Colorado Denver
Campus Box 179, P. O. Box 173364
Denver, CO 80217-3364
Candice.Shelby@ucdenver.edu

Emotions have been analyzed and argued over by philosophers for at least 2500 years, largely with an eye to avoiding them. The reason for this "allergy" to emotions is that they were conceived ofas brute, inexplicable phenomena arising from somewhere other than in the thinking self, and so could only muddle rational thinking. In the 20th century, though, there emerged another view of emotion, articulated by William James, according to which emotions are feelings constituted by perceptions of changes in the body. On this view, one feels fear *because* one's heart races, one's mouth is dry, etc. This view, however, although at least founded in organismic experience, failed to appreciate the motivational aspect of emotion—it seems to be the end, rather than the beginning, of action. Not only that; this view failed to consider that we often are concerned about whether our emotions are *justified*. That we blame someone for unprovoked anger, for example, seems to show that emotions have a rational element to them. That's why we often try to talk our friends out of particular emotions. But judgments still offer no account of the motivational feature of emotions; one might well make a judgment that someone or something is worthy of love or fear, and nevertheless not love or fear that thing or person. The rise of the psychological sciences seemed to offer a more complete, testable view of emotions. Taking its primary inspiration from Darwin's *The Expression of the Emotions in Man and Animals*, this view identified six so-called "basic" emotions,believed to be determined by certain biological markers (hardwiring) which can be discovered in certain "universal"facial expressions. That, too, however, turns out to be untrue, as further critiques of emotion research across varying cultures have revealed assumptions that seem to have obstructed the appreciation of the importance of context in interpreting those facial expressions, and the vastly differing emotion concepts that exist across different peoples. Emotions, according to the results of the most complete studies of them yet available, are constructions, just as much as are our perceptions, memories, and imaginings. Central to their construction are concepts, or concept-like neural firing patterns, involving processes including proprioception, prediction, and prediction error. All of these processes depend, as do most biological processes, not only on biophysics and chemistry, but also, and importantly, on biological coding, at many levels. Since codes are arbitrary sets of rules that are tightly observed but only contingently conserved over varying lengths of time, this analysis of emotion opens the door to both the possibility of re-coding, which provides promise for both psychological healing of such syndromes as PTSD and depression, and for understanding in a better way how our moral reasoning operates, thus suggesting how we might become better at understanding and resolving moral differences among us.

Seventh International Conference in Code Biology - Lužnica (Croatia) 31 August – 4 September 2021

Mathematical regularities in the genetic code: a unifying view based on symmetry and group theory

Diego L. Gonzalez[1], Jean-Yves Boulay[2], Simone Giannerini[3], Greta Goracci[3]
[1]IMM-CNR UOS of Bologna, Italy,
[2]Independent Researcher, Reunion Islands,
[3]Department of Statistical Sciences, UNIBO, Italy
gonzalez@bo.imm.cnr.it

From the elucidation of the genetic code has been clear that the codon to amino acid assignation does not correspond to a random mapping but shows strong mathematical regularities. These regard mainly both, the degeneracy distribution (number of codons coding for the same amino acid), and the physical/chemical properties of amino acids (mainly the number of nucleons). For explaining these regularities different models and theories have been proposed. However, no integrated view has been developed so far. In this paper we propose a unifying framework that connects numerical properties of the degeneracy distribution with numerical properties of the number of nucleons in the coded amino acids. The framework proceeds along both, the development of a model based on the non-power representation of integer numbers, and the analysis of symmetries by means of groups theory. The presented results open the way for a holistic understanding of the genetic code and the elucidation of the biological role of the found mathematical structures.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

What's Wrong with Code Biology

Dušan Gálik
Institute of Philosophy, Slovak Academy of Sciences
Klemensova 19
813 64 Bratislava, Slovak Republic
Dusan.Galik@savba.sk

Code biology is presenting itself as a new science of biology based on code and its essential role in explaining living system. The aim of the paper is to show to what extent the code biology is acceptable as a new approach to investigate living systems and their evolution comparing to "traditional" biology. The main idea of the paper is to show that:

- there are different types of "code" in living systems, that is, there are different meanings of "code" used in explanation of processes in living systems. Or, in another words, there are different types of information, and also different types of "coding", or "codes" in living systems. If there is a way how to save a code approach to biology, than it is necessary to distinguish among these different concepts of code;
- the second problem is to what extent can we call any process in a cell/tissue/organ/organism a process of coding/decoding an information (with the question - what kind of information);
- regarding the different types of codes in living systems there is a significant difference between language code and another types of codes. Unfortunately, this difference is not taken into account in many "code" approaches as seriously as they should be. This is one of the sources of misunderstanding in code biology. If there is a way to save code biology as a specific approach to living nature than it is necessary to express this difference.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

Is Quantum Information Approach Suitable For Modelling The Genetic Code?

Elena Fimmel (Mannheim University of Applied Sciences)
e.fimmel@hs-mannheim.de
Sergey Petoukhov (Russian Academy of Sciences)
spetoukhov@gmail.com

Since the early 1950s, marked by the revolutionary discovery of the double helix structure of DNA by F. Crick and J. Watson, biology has probably been developing most intensively compared to other natural sciences. A lot of empirical data have already been revealed using increasingly sophisticated methods. In spite of a remarkable progress made during the last century, Theoretical Foundations of Biology still remain in a nascent state, which can be compared with the state of Theoretical Physics before Newton. This is the reason why it is necessary to search for a reliable theoretical basis especially in the modelling of the genetic code. This talk aims to show the possibilities of modelling the information content carried by quantum mechanical DNA-molecules by means of the formalism used in quantum informatics. Such modelling would open new options to reveal nature's information patents and to use them.
First of all, a basic concept for modelling the genetic code by means of quantum information formalisms will be presented. It will be shown that the four by the nature as letters for the genetic alphabet chosen nucleotide bases as well as n-nucleotides can be represented in a natural way using their biochemical properties as quantum systems of qubits. In addition, the possibilities to use the operation of the Kronecker (tensor) product, which is very important in the quantum mechanics, to model some properties of the long DNA-sequences will be investigated.

Seventh International Conference in Code Biology - Lužnica (Croatia) 31 August – 4 September 2021

"Double-agents" in oncogenesis – when genes can play on both sides

Hana Owsianková
Department of general linguistics, Palacký University
Olomouc, Czech Republic
hana.owsiankova@upol.cz

As defined by molecular genetics, a gene is a DNA segment that encodes a functional product where in the functional product is meant RNA or its translation product, a protein (polypeptide). The function of these products is directly related to their structure. In this contribution I would like to focus on the structure of genes which encode products with ability to switch their function from the default one to the right opposite, based on cellular context. In particular, my aim is to discuss this structure-function peculiarity observed among human genes whose changes in expression usually cause cancer.

Simply put, cancer is a designation for large group of diseases manifesting by abnormal cell growth and the potential of spreading to the other parts of body – not only environmental but also genetic factors play crucial role in cancer development. In the cells of multicellular organisms we can find both oncogenes and tumor-suppressor genes – oncogenes can cause normal cell to grow out of control and become cancer, tumor-suppressor genes protect them from degenerating into cancer cells. In carcinogenesis, they seem to be contradictory gene groups, yet some genes have dual roles, i.e. display both oncogenic and tumor-suppressor functions under different cellular context (see Soussi – Wiman, 2015; Yang et al., 2007; Yip et al., 2010). Because we are discussing topic concerning genetic code and encoding genetic information, the structure of above described genes will be examined by quantitative methods that primarily fall within the field of linguistics– a field which, besides other things, deals with codes of all types. These methods, e.g. Damerau-Levensthein distance, Bag-of-words model, Manzerath-Altmann law, have been introduced to genetics more recently (see Baixeries et al., 2013; Bolshoy, 2003; Lovato, 2015; Sutton et al, 2014).

References

Baixeries, J. – Hernández-Fernández, A. – Forns, N. – Ferrer-i-Cancho, R. (2013). The Parameters of Menzerath-Altmann Law in Genomes. *Journal of Quantitative Linguistics* 20(2), pp. 94–104.

Bolshoy, A. (2003). DNA sequence analysis linguistic tools: contrast vocabularies, compositional spectra and linguistic complexity. *Applied bioinformatics* 2, pp. 103–112.

Lovato, P. (2015). *Bag of words approaches for Bioinformatic*, Ph.D. thesis, Dept. of Computer Science, University of Verona, series TD-03-15.

Soussi, T. – Wiman, K. G. (2015). TP53: an oncogene in disguise. *Cell Death Differ.* 22, pp. 1239–1249.

Sutton, L. A. et al. (2014). An Entity Evolving into a Community: Defining the Common Ancestor and Evolutionary Trajectory of Chronic Lymphocytic Leukemia Stereotyped Subset #4. *Molecular Medicine* 20(1), pp. 720–728.

Yang, L. – Han, Y. – SuarezSaiz, F. – Minden, M. D. (2007). A tumor suppressor and oncogene: the WT1 story. *Leukemia* 21, pp. 868–876.

Yip, S. C. – Saha, S. – Chernoff, J. (2010). PTP1B: a double agent in metabolism and oncogenesis. *Trends Biochem. Sci.* 35, pp. 442–449.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

Maxwell®: a new classification tool applied to the search for relics of ancient RNAs in modern genomes

Jacques Demongeot[1,2], Joël Gardes[2,3], Hervé Seligmann[2,4], Nicolas Vuillerme[1,2]
[1]Laboratory AGEIS, EA 7407, Faculty of Medicine, University Grenoble Alpes, 38700 La Tronche, France
[2]Labcom CNRS/UGA/Orange®, Faculty of Medicine, University Grenoble Alpes,
[3]OrangeLabs®, 28 Chemin du Vieux Chêne, 38240 Meylan, France
[4] The National Natural History Collections,
The HebrewUniversity of Jerusalem, 91904 Jerusalem, Israel

Maxwell® is an original artificial intelligence platform, which follows an unsupervised learning approach in line developed namely for clustering in genomics. The classification algorithm is based on the Burrows-Wheeler distance, used in compression problems: it identifies the repetitions of common patterns between two data and decreases their Hamming distance (equal to numbers of different elementary patterns) by giving these patterns a compressed expression shorter than the initial expression. An example is given in Figure 1. The reversible Burrows-Wheeler transformation is applied to the anti-palindrome X CAAGCTTG in which the CAAG subsequence is followed by the symmetrized sequence CTTG, in which the Crick transformation was carried out, G (resp. C, A, T) giving C (resp. G, T, A). It consists of writing the 8 circular permutations to the left of the letters of X (left column) and writing, on the right, a column made of the classification in alphabetical order of the left column. The transform of X, denoted BWT (X), consists of the number of the row in the right column containing X, followed by the list of the last letters of this column. We do the same for Y and the Burrows-Wheeler distance between X and Y is equal to the Hamming distance between BWT (X) and BWT (Y). In the example of Figure 1, this distance has been reduced by 2, compared to the Hamming distance between X and Y (from 6 to 4), thus showing the advantage of the method for bringing structures containing patterns with the same characteristics (here the anti-palidromic character). If the data are genomic, other characteristics can be detected by the Burrows-Wheeler transformation, such as insertion, repetition, inversion, translocation, deletion, symmetrization, palindromization, Seligmann swinger transformations, etc. If they are common between the data like the pentamers issued from ancient RNA rings selected from genetic code constraints [1-7], their presence causes a reduction in their distance, favoring the assignment to the same class.

References

[1] J. Demongeot, H. Seligmann. Theoretical minimal RNA rings recapitulate the order of the genetic code's codon-amino acid assignments. JTB, 471, 108-116 (2019).
[2] J. Demongeot, H. Seligmann. Spontaneous evolution of circular codes in theoretical minimal RNA rings. Gene 705, 95-102 (2019).
[3] J. Demongeot, H. Seligmann. More pieces of ancient than recent theoretical minimal proto-tRNA-like RNA rings in genes coding for tRNAsynthetases. J. Mol. Evol. 87, 152-174 (2019).
[4] J. Demongeot & H. Seligmann. Bias for 3'-dominant codon directional asymmetry in theoretical minimal RNA rings.J. Computational Biology 26, 1003-1012 (2019).
[5] J. Demongeot, V. Norris. Emergence of a “Cyclosome” in a Primitive Network Capable of Building “Infinite” Proteins. Life 9, 51 (2019).
[6] J. Demongeot, H. Seligmann. Theoretical minimal RNA rings designed according to coding constraints mimick deamination gradients. Die Naturwissenschaften / The Science of Nature 106, 44(2019).
[7] H. Seligmann, J. Demongeot. Codon directional asymmetry suggests swapped 1st and 2nd codon positions.IJMS 21, 347 (2020).

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

Junguian Archetypes: metaphors or biological entities?

João Carlos Major
International Academy of Analytical Psychology - IAAP|pt
Braga - Portugal
jcmajor@iaap.pt

The reality of 'archetypes', that is, archaic behavioral patterns that are present in the human species regardless of the underlying culture and highly conserved in evolution, is a well-documented theme in analytical psychology. They are, in fact, the basis of analytical psychology in theoretical and clinical terms; which brings analytical psychology closer to ethology and evolutionary psychology, which it pioneered. But what are they? And, if they really exist, where are the archetypes? C.G. Jung has suggested the term *objective psyche* for that *a priori* (phylogenetic) part of the psyche which elicits behaviors based on emotionally competent stimuli. This objective psyche manifested in emotions and drive impulses (archetypal behaviors). In fact, any animal has a modeling system that represents some proprieties of the external world – the Uexküll's *Umwelt* (1909) or Carl Jung's *Psychological Archetypes* (1921) – and determines the behavioral response to them. Hence, the ego-centered, subjective consciousness is a partial rather than a complete manifestation of the psyche, as presented by William James. In this paper we will defend the hypothesis that archetypes – understood as a psycho-physical realities, a set of memories inherited from the species, or a set of codified rules – are real biological entities or codes, that have been highly conserved in evolution, remnants of behaviors that once allowed the survival of what became homo sapiens, probably arranged in the form of 'neuronal codes' (the rules by which the animal sense organs transform the incoming signals into neural states). In this sense, they are presented as a concept correlated to 'instinct' or 'primary emotions', and must be taken into account in psychological studies and, fundamentally, in clinical practice in psychology. In fact, it's time for psychology to reopen itself to biology, overcome the 'main stream' cognitive-behavioral perspective whose socio-historical roots based on the idea of *tabula rasa* still overshadow contemporary psychology, hampering its development as a science.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

Minerobiolization: Chemobrionics and biomineralization at the origin of life

Julyan Cartwright[1]

[1]Instituto Carlos I de Física Teórica y Computacional
Universidad de Granada, 18071 Granada, Spain
julyan@onsager.ugr.es

If, as seems likely, life developed in hydrothermal vents, it necessarily first used mineral membranes within the vents. The first proto-cells must have learnt to manipulate the mineral membranes that formed their compartments in order to control their metabolism. There must have occurred a biological takeover of the self-assembled mineral structures in the first proto-cells, with the incorporation of proto-biological molecules within the mineral membranes to alter their properties for life's purposes, so that, for example, passive osmosis in a mineral membrane gradually became active chemiosmosis in a proto-cell membrane. For this biological takeover of a mineral system we coin the term minerobiolization, in contradistinction to the usual biomineralization in which the biology controls and assembles the mineral.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

On the asymmetrical relation between linearity and non-linearity

Ľudmila Lacková
Department of General Linguistics
Palacký University, Olomouc, Czech Republic
ludmila.lac@gmail.com

According to Barbieri, organic codes differ from physical laws by the notion of arbitrariness. Organic codes differ from physical laws but differ also from engineering codes in that the biological codes transform a sequence of a given alphabet to a sequence from another alphabet, whereas engineering codes transform only from one and the same alphabet (Barbieri 2018). Barbieri's definition of biological codes is clear, nonetheless there are some characteristics of biological codes that were not discussed in detail from Code Biology perspective. So far, what is generally understood by a code, whatever its type, is connected to sequences. Yet not every code is to be sequential and organic codes also code for structures that are not sequential, but of another order. In my contribution, I would like to analyze the non-sequential (non-linear) character of folded protein structures in relation with sequences coding for them.
The relation between linear sequence and non-linear structure is characteristic both for protein folding and human language (syntax for instance). In recent years, analogies between protein folding and grammars of natural language have been suggested (Gimona 2006, Kister 2015). From a linguistic standpoint, I propose that the analogy between natural language and the protein folding process is linked by only one special design feature. This feature can be termed the asymmetrical relation between linearity and non-linearity. Both grammars of natural language and protein studies deal with linearity problems. This is evident in the tension between sound and meaning in natural language and the tension between the peptide chain and protein shape in protein folding.

References

Barbieri, M. (2018). What is code biology? *BioSystems* 164 (2018) 1–10.
Gimona, M. 2006 "Protein linguistics – a grammar for modular protein assembly?", Nature Reviews Molecular Cell Biology, 7 (Jan. 2006), 68 EP.
Kister, A. 2015 "Amino acid distribution rules predict protein fold: protein grammar for beta-strand sandwich-like structures", Biomolecules, 5, 1 (Jan. 2015), 41-59.

Seventh International Conference in Code Biology - Lužnica (Croatia) 31 August – 4 September 2021

Applying Mechanistic Explanation in Code Biology

Lukáš Zámečník,
Palacký University in Olomouc
lukas.zamencnik@upol.cz

Contemporary philosophers of science suggest several new models of scientific explanation that are suitable for application in code biology – mechanistic (Craver, Darden 2013), design (Eck, Mennes 2016), or various variants of non-causal explanations (Reutlinger, Saatsi 2018). The paper analyzes the mechanistic model of explanation in its three different variants – when mechanisms are producing, underlying, and maintaining phenomena (e.g. Glennan 2016). The goal is the application of these variants of explanation in several cases of extended mechanisms built in code biology (Barbieri 2019, 2015).The main goal is the explication of the role of information and code in extended mechanisms in code biology – in contrast with the simplifying view (that "life is chemistry") present in some mechanist philosophers (e.g. Glennan 2017).

References:

Barbieri, M., 2019. Evolution of the Genetic Code: The ambiguity-reduction theory. BioSystems. 185, 1-7.
Barbieri, M.,2015. Code Biology: A New Science of Life. Springer, Dordrecht.
Craver, C.F., Darden, L.,2013. In Search of Mechanisms: Discovery across the Life Sciences. The University of Chicago Press, Chicago.
Eck, D. v., Mennes, J.,2016. Design Explanation and Idealization. Erkenntnis. 81, 1051-1071.
Glennan, S., 2016. Mechanisms and Mechanical Philosophy, in: Humphreys, P. (Ed.) The Oxford Handbook of the Philosophy of Science. Oxford University Press, Oxford, 796-816.
Glennan, S., 2017. The New Mechanical Philosophy. Oxford University Press, Oxford.
Reutlinger, A., Saatsi, J.,2018. Explanation Beyond Causation. Oxford University Press, Oxford.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

Circular tessera codes and their role in the evolution of the genetic code

Lutz Strüngmann
(joint work with Elena Fimmel and Martin Starman)
Institute for Mathematical Biology,Faculty of Computer Science,
Mannheim University of Applied Sciences, 68163 Mannheim, Germany
l.struengmann@hs-mannheim.de

The origin of the modern genetic code and the evolutionary process that has turned an ancestor code into its present form raise many questions [1]. In [2] Gonzalez, Giannerini and Rosa developed a theory of a tessera code, based on tetranucleotides, that can explain the degeneracy of the modern vertebrate mitochondrial code. They also explained how these tesserae could have been read by tRNA adaptors that are symmetric and could read both DNA strands in both directions. From an evolutionary point of view this might imply that the genetic code evolved by first counting in powers of 2, i.e. passing from one nucleotide to dinucleotides and then to tetranucleotides, and after that by an information reduction to trinucleotides (see e.g. [3]). The role of dinucleotides in evolution is supported by several works, e.g. [4] where it was hypothesized that two independent sets of dinucleotides (prefix and suffix doublets) were at the beginning of the genetic code or by [5] where the possibility of reversible ancient tRNAs was claimed. Those would have been able to read dinucleotides in both directions. Moreover, in the 1990s, so-called circular codes were discovered in the genetic code, which seem to ensure the maintenance of a correct reading-frame during the translation process. A corresponding theory has been developed since then providing another evolutionary model for the genetic code. This talk will describe the attempt to combine the two concepts, the tesserae and the circular codes, and see if they could benefit from each other. It turns out that the two concepts fit together nicely. In particular, we will give several properties, construction principles and graph-theoretical characterisations of circular tessera codes.

[1] Fimmel E., Starman, M., and Strüngmann L.: Circular tesserae codes and their role in the evolution of the genetic code, submitted (2019).
[2] Gonzalez D.L., Giannerini S., Rosa R.: {\it On the origin of degeneracy in the genetic code.}, in Interface Focus 9: 20190038. http://dx.doi.org/10.1098/rsfs.2019.0038 (2019)
[3] Baranov P.V., Venin M., and Provan G. : Codon Size Reduction as the Origin of the Triplet Genetic Code,
PLOS ONE 4(5),2009: e5708. doi:10.1371
[4] Wu HL, Bagby S , and van den Elsen JM: Evolution of the genetic triplet code via two types of doublet codons, J Mol Evol. 61(1), 2005, 54-64.
[5] Wilhelm T, and Nikolajewa S.: A new classification scheme of the genetic code, J. Mol. Evol., 59(5), 2004, 598-605.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

Metaphors of Identity, Contact and Connection Literature and Cell Biology during the 20th Century

Marcella Faria
Dactyl Foundation for the Arts and Humanities
64 Grand Street, New York, NY 10013 USA
almeidapradomarcella@gmail.com

The widespread practice of indexing research articles with key words and analyzing data trough semantic parameters are symptomatic of the importance of "words" to contemporary science. One of the most iconic terms in biology is "cell", it originally comes from the Latin *cella*, *i*.e. small room. As it initially referred to "the place of life", a mere envelope or cavity to be filled by living essence. Nevertheless, it became progressively clear that not only the cells are at the origin of all organism's development (Virchow's phrase *"Omnis cellula e cellula"*), but they are systems. Therefore, the term "cell" has dynamically changed from a strictly structural perspective into a rather functional one. The emerging picture resulting from our previous studies points to the acquisition of distinctive properties by the cells through evolution. These features can be summarized into discrete sets of key words: 1) identity, boundary conditions, membrane, autonomy, cloning, closure, invasion, starvation, isolation; 2) recognition, contact, pattern, migration, polarization, development, transformation, differentiation; 3) communication, network, connection, memory, cognition, activation. The sets of properties acquired by "cells" can be analyzed in contexts as diverse as the phylogenetic evolution of species and the embryonic development of individuals. This list of scientific terms parallels the progression of accumulated knowledge in Cell Biology, but it is also quite straightforward that it has correlates in other domains. We will be interested in identifying potential overlaps between "cell biology terms", historical facts and literary pieces. The inspiration for this kind of comparative research comes from the work of three iconic figures, who are coincidentally women and have a double training in science and literature, namely: Laura Otis, Donna Haraway and Joan Anderson. It is important to notice that the concepts in cell biology are organized into three discrete categories. Progression through this three-step process always match the coming into being of new organic codes, or the complexification of already existing ones. Explaining major phylogenetic changes, the so called macroevolutionary transitions. At the same time, they recapitulate at the historical time scale of fictional narratives and political changes, the changes in metaphorical representations for identity, contact and communication through the last Century.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

Unity, Diversity and Pluralism in Code Biology

Marcello Barbieri
Dipartimento di Morfologia ed Embriologia
Via Fossato di Mortara 64a, 44121 Ferrara, Italy
brr@unife.it

In the past few years there have been various proposals that have extended the original theory of Code Biology. In order to discuss these extended theories, however, it is necessary to start from a summary of the original theory, and this can be done by summarizing the four points that make this theory different from four major theoretical frameworks:

[1] The original theory of Code Biology is different from the *Modern Synthesis* for two reasons. The first is the idea that evolution took place by natural selection and by natural conventions and these mechanisms are fundamentally different because natural selection is based on copying and natural conventions are based on coding. The second is the idea that the cell is not a duality of genotype and phenotype but a trinity of genotype, phenotype and ribotype, where the ribotype is the ribonucleoprotein system of the cell that functions as the codemaker of the genetic code (Barbieri 1981, 1985, 2003).

[2] The original theory of Code Biology maintains that the basic process of life is not autopoiesis but codepoiesis. Autopoiesis requires biological specificity and specificity comes from the genetic code, so the ancestral systems that came before that code could not have been autopoietic systems. Those ancestral systems, on the other hand, were engaged in the evolution of the genetic code and were therefore codepoietic systems. It is a fact, furthermore, that the eukaryotes increased their complexity during the history of life and their evolution was accompanied by new codes, which suggests a deep link between codepoiesis and complexity, i.e., between the origin of new codes and the origin of the great novelties of macroevolution (Barbieri 2012, 2015, 2017).

[3] The original theory of Code Biology is different from *Biosemiotics* because it claims that the Peircean processes of 'interpretation' and 'abduction' take place in the brain but not in the cell (Barbieri 2014, 2018).

[4] The original theory of Code Biology is different from the *Relational Biology* of Robert Rosen because it assumes that the process of 'anticipation' takes place in the brain but not in the cell (Barbieri 2019).

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

Translation in the Light of Origins of Life, Enzyme Evolution and Definition of Life

Marko Vitas
Laze priBorovnici 38, 1353 Borovnica, Slovenia
vitas.marko83@gmail.com
and
Andrej Dobovišek
University of Maribor, Faculty of Natural Sciences and Mathematics,
Koroška cesta 160, 2000 Maribor, Slovenia

A new definition of life has been formulated by modifying and extending NASA's working definition of life. The essence of life can be defined as follows: *Life is a far from equilibrium self-maintaining chemical system capable of processing, transforming and accumulating information acquired from the environment*. The new definition includes a thermodynamical aspect of life as a far from the equilibrium system and considers the flow of information from the environment to the living system. In our derivation of the definition of life we have assumed the hypothesis, that during the emergence of life evolution had to first involve autocatalytic systems that only subsequently acquired the capacity of genetic heredity. It is possible, in this sense, to view the ribosome as a digital-to-analogue information converter. Moreover, on the same basis, we propose and discuss possible mechanisms, basic aspects of the emergence and subsequent molecular evolution of translation and ribosomes, as well as enzymes as we know them today. In this sense, it was hypothesised that primordial RNAs were devoid of any repository role for discrete digital information. Furthermore, we propose the acquisition of the primary protein structure as a first step in digitalisation of information contained in biological system, achieved by accretion. Cytochromes P450 have been proposed as a plausible model for the evolutionary development of protein catalysers. The proposed mechanism is based on the abilities and tendencies of short RNA and polypeptides to fold and to catalyse biochemical reactions. The proposed mechanism is in concordance with the hypothesis of a possible chemical co-evolution of RNA and proteins in the origin of the genetic code or even more generally at the early evolution of life on Earth. The hypothesis that early polypeptides were folding on the RNA-scaffold is also considered and mutualism in molecular evolutionary development of RNA and peptides is favoured. Explaining the emergence of life is perhaps central and the most challenging question in modern science. Within this area of research, the emergence and evolution of the genetic code is supposed to be a critical transition in the evolution of modern organisms. Therefore, studying origins of life, origins of genetic code or processing of any kind of biological information may be of interest for astrobiology or for efforts to produce synthetic or artificial life, and it furthermore may also have implications in the cognitive and computer sciences. The results of such investigations and definitions of life might also shift current biological paradigms and might also have a momentous significance for modern philosophy in understanding our place in universe.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

Circular code motif lengths analysis in biological coding sequences

Markus Gumbel
Competence Center for Algorithmic and Mathematical Methods in Biology, Biotechnology and Medicine, Mannheim University of Applied Sciences
m.gumbel@hs-mannheim.de

It is theoretically possible to detect a frame shift at the ribosome during the translation process if the RNA sequence would consist only of codons from circular codes (overview in [1]). There exist many circular codes but in biological sequences the X code discovered by Arques and Michel [2] is most important. Coding sequences (CDS) contain codons of the X code more frequently than any other codons. A frameshift is recognized latest after reading two to three codons in the wrong frame. Any circular code can only have 20 codons at maximum which is problematic since coding sequences typically contain all 64 codons. A consecutive sequence of codons from a circular code is called a motif. The longer such a motif is the higher is the chance to detect a frame shift. Clearly, there is a high chance that a coding sequence is interrupted many times by codons not part of a circular code. Thus, a frame shift cannot always be detected in these coding sequences. However, this might have been the case in very old sequences of ancient organisms – maybe in species as old as the genetic code itself. If the circular code's ability to detect frame shifts was (or still is) applied one would expect that the lengths of motifs in current organisms are longer (on average) than expected by chance. In this presentation we analyze the motif lengths in biological coding sequences of different species using a set of relevant circular codes (including X code).

[1] Elena Fimmel and Lutz Strüngmann: Mathematical fundamentals for the noise immunity of the genetic code. BioSystems, 164: 186–198, 2018
[2] Didier G Arquès and Christian J Michel: A complementary circular code in the protein coding genes. Journal of theoretical biology, 182(1):45–58, 1996.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

High optimization for genetic code

Massimo Di Giulio
The Ionian School, Genetic Code and tRNA Origin Laboratory
Via Roma 19, 67030 Alfedena (L'Aquila), Italy
and
Institute of Biosciences and Bioresources, National Research Council,
Via P. Castellino 111, 80131 Naples, Italy
massimodigiulio1952@gmail.com

We (Caldararo and Di Giulio, 2020) have analysed a model for the origin of the genetic code that takes into account both the physicochemical properties of amino acids and their biosynthetic relationships. In particular, we considered the set of amino acid permutation codes subject to biosynthetic constraints. We analysed the level of optimization achieved by the genetic code in this constrained set, considering 545 physicochemical or biological properties of amino acids. Our conclusion is that overall the optimization of the structure of the genetic code is close to that of absolute maximum. We believe that these observations are better explained by the coevolution theory of the origin of the genetic code than by physicochemical theories.

Seventh International Conference in Code Biology - Lužnica (Croatia) 31 August – 4 September 2021

Chemical Etiology of the Canonical Amino Acid Repertoire: The Alanine World Model

Nediljko Budisa
Berlin Institute of Technology/TU Berlin, Department of Chemistry, Biocatalysis Group,
Müller-Breslau-Straße 10, D-10623 Berlin, Germany.
University of Manitoba, Department of Chemistry & Microbiology
Chair of Chemical Synthetic Biology, 144 Dysart Rd, R3T 2N2 Winnipeg, MB, Canada
budisan@umanitoba.ca

Life on earth is a unity owing to the existence of the universal genetic code, i.e. as the genetic code for all organisms is basically the same - all living things use the same "genetic language". This biological framework allows the translation of the genetic message written in DNA into life-sustaining proteins build with 20 canonical amino acids. Why ‘only’ 20?According to Woese's early suggestion, primitive cells or protocells began with a completely random, highly ambiguous set of codon assignments to amino acids with very inaccurate translation. The process of expanding the amino acid repertoire ended with the "frozen accident". Therefore, the answer to the question why only 20 amino acids are the standard repertoire of the universal genetic code is an evolutionary one.

Recently, we proposed Alanine World model that explains the choice of amino acids monomers in the genetic code repertoire from the viewpoint of peptide and protein chemistry. The core of our considerations is the selection of secondary structural elements for the construction of the protein scaffolds and the subsequent life body plans (forms, morphology). This also determines the choice of amino acid monomers. Dominant secondary structures in life as we know it are α-helices and ß-sheets, which are mainly made up of alanine derivatives (from a chemical point of view; that is why we coined the term "Alanine World"). In this model, the selection of monomers (amino acids) for such secondary structural elements is rather limited. For example, chemical propensities for α-helix or ß-sheet determine exactly which side chains of alanine derivatives are suitable and which are not.

The Alanine World model, together with the RNA world hypothesis (along with Hartman-Smith model) and the Co-evolution theory, offers a plausible scenario for the chemical etiology of the amino acid repertoire in the genetic code. In addition, this is also a working model and a good starting point for the further development of code engineering towards a truly synthetic life with novel genetic codes. The Alanine World model, also supports Code Biology's main postulate, i.e. that associations between amino acids and codons are arbitrary (this however, only applies in the initial stages of establishing the genetic code).

Finally, the Alanine World hypothesis should be of great importance for Astrobiology because it makes it possible to predict a scenario for the alien life with genetic codes having completely different amino acid sets (supplied by the orthogonal metabolism) with corresponding secondary structure scaffolds. The alien life forms based on these structures and processes may be better suited for living habitats that are fundamentally different from the one on Earth.

Kubyshkin, V., Budisa, N: The Alanine World Model for the Development of the Amino Acid Repertoire in Protein Biosynthesis. *Int. J. Mol. Sci.* **2019**,*20(21)*, e5507.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

Complementary Peptide Interaction: Applications and Perspectives

Nikola Štambuk
Ruđer Bošković Institute, Zagreb, Croatia
stambuk@irb.hr

The concept of complementary peptide interaction was developed in the early 1980s by Root-Bernstein, Biro, Blalock, Mekler, Siemion, and others. It was deduced from empirical observations that peptides consisting of complementary (sense and antisense) amino acids interact with higher probability and affinity than the randomly selected ones. This phenomenon is closely related to the structure of the standard genetic code table, and at the same time, is unrelated to the direction of its codon sequence translation. Consequently, the natural genetic coding algorithm for sense and antisense peptide interactions combines elements of amino acid physico-chemical properties, stereochemical interactions, and bidirectional transcription. The second codon base specifies the physicochemical properties of the amino acids. Therefore, it is not surprising that diverse amino acid properties—like hydrophobicity, hydrophilicity, lipophilicity, and molecular descriptors of contact potential (Miyazawa-Jernigan), hydrophobic moment, and intrinsic disorder—follow the identical sense and antisense complementarity clustering scheme that is associated with molecular interaction at the peptide level (of ≥ 4 amino acids).
Antisense peptide technology is based on a heuristic algorithm for rational peptide design of the interacting ligand–acceptor (receptor) sequences specified by the complementary codons. Heuristic methods reduce solution space by focusing on results based on the reduced set of criteria—in this case, complementarity rules defined by the standard genetic code table. Additionally, it was also shown that modern computational methods enable a new approach to the studies of sense and antisense peptide interactions. The benefits of antisense peptide technology outweigh the costs of random peptide screening and could lead to considerable savings in time and resources, especially if combined with other computational and immunochemical methods. The applicability of antisense peptide technology was confirmed recently for immunoassays and immunohistochemistry. This opens a perspective for the development of a new class of efficient biomedical procedures based on short peptide technology.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

A code-based framework of human regulome

O. Paredes, R. Romo-Vázquez, J. A. Morales
Departamento de Ciencias Computacionales, CUCEI,
Universidad de Guadalajara, México
omar.paredes@alumnos.udg.mx

Organic codes are mapping rules of biological molecules, states or entities in organisms. Adaptors translate these rules into copying/coding mechanisms. However, some biological codes could code for the same set of features with converging biological outcomes. The genome is a unit that contains the information of an organism, but such units may have different phenotypical states coded within it (1). The regulome is the regulatory network that determines such phenotypic states, and these are carried out by regulons. Therefore, a regulon is a microcosmos where epigenomic phenomena and regulatory binding sites within the genome controls genetic co-expression networks. In this work, we propose a code-based model of the human regulome where the histone code and the genomic codes (mainly the regulatory codes), constitute regulons that outline cell lines. This 'code of codes' reveals that organic codes are not entities enclosed in single biological layers but lead to a continuum of arbitrary mapping rules through biological layers in the organisms from the ancient, immutable genetic code to the novel, evolving, cultural codes.
(1) R. Gordon, N. K. Gordon, *BioSystems*. 184, 104013 (2019)

Seventh International Conference in Code Biology - Lužnica (Croatia) 31 August – 4 September 2021

Relational Model of the Standard Genetic Code

Paško Konjevoda and Nikola Štambuk
Ruđer Bošković Institute,
Bijenička cesta 54, 10000 Zagreb, Croatia
pkonjev@irb.hr (P.K.) *and* stambuk@irb.hr (N.Š.)

The genetic code is a set of rules that establish a mapping between triplets in messenger RNA and amino acids in proteins. The most common way to display these rules is the Standard Genetic Code (SGC) table. The position of each amino acid in the SGC table is defined by bases of the related triplet: the first base defines rows, the second base defines columns, and the third base defines subrows. There are four subrows in each row, sixteen in total. Consequently, the fact that codon position within the SGC table is specified by three bases suggests that the genetic code is a more complex structure than a simple 2D table. Over the last half of the century numerous alternative models have been applied to resolve the SGC conundrum. This paper takes an alternative approach, based on the relational data model. In 1970, Edgar F. Codd, a computer scientist working for IBM, published a highly influential paper which introduced the theoretical concept of the relational model RM (Commun. ACM, 13:377−387, 1970). Explained in a condensed way, the relational model proposes distributed storage of data into a collection of tables (which this model calls relations), that which can be connected by shared communality. Basic elements of the SGC table are rows (called records or tuples), and columns (called fields or attributes). The SGC table, according to the relational data model, represents the so called unnormalized form of a table. In other words, subrows break atomicity, the rule that requires that only one piece of data is allowed at a point where a column crosses a row. An unnormalized table can be decomposed or divided into more than one table, using a set of rules called normal forms. Tables produced in this way are called normalized forms. There are six levels of normalization, and the first three levels are sufficient in most cases. Using normalization rules it is possible to subdivide the SGC table into four tables. The rows and columns of single tables are defined by the first and second base and individual tables by the third codon base. The result of this model is an approach to managing genetic code data, represented in terms of tuples and grouped into relations, with table structure and language consistent with first-order (predicate) logic. Biological interpretation of the relational model suggests that the primary force driving then evolution of the genetic code was incorporation of new amino acids into a coding system. However, additional amino acids were not incrementally added to a coding system consisting of 64 triplets. The precursor of SGC was a population of several simpler coding systems consisting of 16 doublets, and following this different amino acids were added to the coding systems. The third base was probably present, but without the coding function, as a simple stabilizing anchor. One of the adaptations was a synchronization of coding systems by physicochemical properties¾identical or maximally similar amino acids were positioned at the same position in different codes. Coding systems structured in this way enabled a synthesis of statistical proteins, i.e. proteins that shared a common core, with some amino acids that slightly oscillated in physicochemical properties. The relational model explains that the final step in the development of SGC was the adoption of coding function by the third base, which in SGC constitutes an informational/functional unit with the first base, despite their different physical location in a triplet. This enabled the synthesis of specific proteins without ambiguity, in accordance with the concept of ambiguity reduction and five phases of the general model on the origin of biological codes by Marcello Barbieri (BioSystems 181:11−19, 2019).

Seventh International Conference in Code Biology - Lužnica (Croatia) 31 August – 4 September 2021

Theoretical model of genetic code structure evolution

Paweł Błażej, Małgorzata Wnętrzak, Paweł Mackiewicz
Department of Bioinformatics and Genomics,
Faculty of Biotechnology,University of Wroc ław,
ul. Joliot-Curie 14a, Wrocław, Poland
blazej.pawel@gmail.com

In this work, we present a theoretical model of genetic code evolution where potential genetic codes evolved from system of ambiguous codon assignments to the set of rules with reduced level of uncertainty. We are especially interested in testing some factors which would be responsible for the compact structure of the standard genetic code. In this context, we consider three types of codon reading systems creating codon blocks assigned to respective amino acid. To do so, we run simulations where mistranslations can act only on fixed positions in codon. This process starts from randomly generated genetic codes with ambiguous codon assignments and they are selected to improve coding ambiguity and their robustness against misreading. The results indicate that the evolution of genetic code carried out under restrictions similar to those which are observed in the standard genetic code is responsible for finding solutions with the best quality.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

The algebraic features of the genetic code and of inherited bio-structures

Sergey V. Petoukhov
Mechanical Engineering Research Institute of Russian Academy of Sciences
Russia, Moscow, MalyiKharitonievskypereulok, 4.
spetoukhov@gmail.com

According to the founders of quantum mechanics P. Jordan and E. Schrödinger, the main difference between living and inanimate objects is the dictatorial influence of genetic molecules on the whole living organism (in contrast, inanimate objects are controlled by the average random movement of their millions of particles) [1]. The presentation is devoted to an algebraic approach in the field of code biology, which deals with algebraic modeling of the structural organization of DNA/RNA alphabets and of inherited biological macro-phenomena bearing the stamp of these alphabetical structures. This scientific direction is closely related to the "grammar of biology". It initially focuses on the binary-oppositional alphabetic structure of DNA and RNA molecules, which leads to tensor families of special matrices called genetic [2]. This family, in particular, includes a (8*8)-matrix of 64 triplets, in which encoded 20 amino acids and stop-codons of protein synthesis are placed in an algebraically regular way associated with hypercomplex numbers [2, 3]. This testifies to the fact that the genetic code is not a simple matching of one set of elements to another set like as in a phone book, in which phone numbers encode the names of people; but that the genetic code is inherently an algebraic code, similar to those algebraic codes that are used for noise-resistant transmission of information in space communication. Given these and other findings, the author believes that living organisms are algebraically encoded essences. The lecture presents argumentations for this idea, showing the following inherited biological structures that are effectively modeled by 2^n-dimensional hyperbolic (double) numbers: laws of phyllotaxis; the spatial-temporal organization of locomotion control in animals and humans; the main psychophysical law of Weber-Fechner; the Mendel's law of inheritance of traits; the basic law of the population genetics by Hardy-Weinberg related to Newton's binomial; rules for the percentage composition of hydrogen bonds 3 and 2 in the DNA of various genomes; similar rules for the percentage composition of two phonetic groups of letters of Russian alphabet in long texts by Tolstoy, Dostoevsky, Pushkin, etc. [3]. The presented examples of interrelated algebraic models of inherited bio-structures make one recall the results of G. Mendel, who showed that, despite the exceptional heterogeneity of the structure of living bodies consisting of myriads of different molecules, the transmission of hereditary traits obeys certain algebraic relations. Additional author's data testify that presented series of algebraic bio-phenomena can be related with resonance interrelations in oscillatory biosystems with many degrees of freedom (the author's conception of multi-resonance genetics and morphogenesis [4]).

References

1. McFadden J., Al-Khalili J. (2018). The origins of quantum biology. *Proceedings of the Royal Society A*: *Mathematical, Physical and Engineering Sciences*. 12.12.
2. Petoukhov S.V., He M.(2009). *Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics*. Hershey, USA: IGI Global.
3. Petoukhov, S.V. (2019). The Genetic Code, Algebraic Codes and Double Numbers. *Preprints 2019*, 2019110301 (doi: 10.20944/preprints201911.0301.v1).
4. Petoukhov S.V. (2016 January). The system-resonance approach in modeling genetic structures. *Biosystems*, v. 139, pp.1-11.

Seventh International Conference in Code Biology - Lužnica (Croatia) 31 August – 4 September 2021

A role for circular code properties in translation

Simone Giannerini
Department of Statistical Sciences, University of Bologna, Bologna, 40126, Italy
simone.giannerini@unibo.it

Alberto Danielli
Department of Pharmacy and Biotechnology, University of Bologna, Bologna, 40126, Italy
alberto.danielli@unibo.it

Diego Luis Gonzalez
Department of Statistical Sciences, University of Bologna, Bologna, 40126, Italy
Institute for Microelectronics and Microsystems - Bologna Unit, CNR, Bologna, 40129, Italy
gonzalez@bo.imm.cnr.it

Greta Goracci
Department of Statistical Sciences, University of Bologna, Bologna, 40126, Italy
greta.goracci@unibo.it

Circular codes represent a form of coding allowing detection/correction of frame-shift errors. Building on recent theoretical advances on circular codes, we provide evidence that protein coding sequences exhibit in-frame circular code marks, that are absent in introns and are intimately linked to the keto-amino transformation of codon bases. These properties strongly correlate with translation speed, codon influence and protein expression levels. Strikingly, circular code marks are absent at the beginning of coding sequences, but stably occur 40 codons after the initiator codon, hinting at the translation elongation process. Finally, we use the lens of circular codes to show that codon influence on translation correlates with the strong-weak dichotomy of the first two bases of the codon. The results provide promising universal tools for sequence indicators and sequence optimization for bioinformatics and biotechnological applications, and can shed light on the molecular mechanisms behind the decoding process.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

On the Grammar and Grammatical Categories of the Genetic Code

Suren Zolyan
Immanuel Kant Baltic Federal University, Kaliningrad, Russia;
Institute of Philosophy, National Academy of Sciences, Yerevan, Armenia
surenzolyan@gmail.com

1. We re-address the possibilities of the description of the genetic code as a sign system: thus is - to identify its grammar and vocabulary. The main distinctions of our approach from the previously suggested grammars are:
a) We make a differentiation between "language" and "speech," between a system of abstract syntagmatic and paradigmatic relations, and its actual manifestation. It can also be represented as a dichotomy of biochemical substance and semiotic form.
b) Instead of linear context-free linguistic models, we suggest using some form of categorial grammar, where items are considered to be context-dependent variables and, simultaneously, context-forming operators. (as it was foreseen by R. Jakobson (Jakobson 1970: 439).
c) As a minimal unit of the alphabet, we consider distinctive features of nuclei acids: 1) the number of bonds – 2 vs. 3; 2) type of base – purine vs. pyrimidine.
d) The substantial distinction is drawn between units of the vocabulary (nuclei acids) and the categories of grammar: the empty positions within triplets (first, second, third), each of them is endowed by its codon-forming functions regardless of which nucleotide it is filled with.
2. The distinction between vocabulary (nucleotides) and categories of grammar (empty positions within triplet) allows to identify the formation rules for the significant units of the genetic code (doublets and triplets) and explicate their compositional semantics (correspondence rules between codons and amino acids). The principle of context-sensitivity allows describing cases when biochemically same sequence of nucleotides, depending on their location, acquires a different meaning and performs a different function.
3. The evolutionary perspective reversed the functional relationship between the first and second positions, but this does not affect the principle itself: one of the positions determines a class of encoded amino acids, the another specifies a member of the class. The complication of the life forms leads to the appearance of new amino acids and new principles of their structuring. The doublet code is transformed into the triplet code. The functional inequality of the third position correlates with its late appearance.

Seventh International Conference in Code Biology - Lužnica (Croatia) 31 August – 4 September 2021

Surpassing the Turing Test: What is the difference between art and science, biological intelligence and artificial intelligence, qualitative and quantitative approaches?

Victoria N Alexander
Dactyl Foundation, New York, USA
Fulbright Scholar, Digital Humanities Deparment
ITMO University, St Petersburg, Russia
alexander@dactyl.org

The goal of a quantitative approach to copying Rembrandt's style of portraiture would be to create a new but typical painting, such as the one produced by the deep-learning algorithm used in the Next Rembrandt experiment (https://www.nextrembrandt.com). The goal of a qualitative approach to copying Rembrandt's style would be to create a unique and original painting such as *Old Woman Cutting her Nails* (1655-60), a painting previously attributed to Rembrandt that is now considered to be the masterful work of one of his students. We can say that Rembrandt's student learned the Rembrandt algorithm for painting portraits. *Old Woman Cutting her Nails* attracts many museum visitors, whereas more typical paintings by Rembrandt, such as those used to train the algorithm in the Next Rembrandt experiment, do not attract the interest of as many visitors or Rembrandt scholars. While the AI-produced painting may pass the Turing Test, *Old Woman Cutting her Nails* surpasses it. This may lead us to ask, What is the difference between the way a biological organism—like one of Rembrandt's students—develops algorithms and the way that machines (currently) do? This talk will explore the biological mechanisms that are employed when new procedures, new biological conventions and codes, new self-confirming habits, are discovered that repeat the old procedures with a significant difference to create a truly original work of art.

My talk will draw upon Alan Turing's late work in morphogenesis and the field of code biology to investigate the sometimes formulaic, sometimes original, mechanisms of biological processes. I will look at biological semiosis as a means-end (i.e. self-reinforcing) process, whose intermediary steps follow arbitrary rules and therefore can function like digital encryption or symbolic codes. At the same time I will examine how these intermediary steps can also be analog codes whose functions are constrained by their physical qualities, such as shape or position. For example, molecular adaptors are shaped such that they can link between first and second messengers. Therefore substitutions, by any similarly disposed tool, are possible, opening the door for creative departures in the program, original works of art or biological adaptations.

Seventh International Conference in Code Biology - Lužnica (Croatia) 31 August – 4 September 2021

Menzerath-Altmann Law on Secondary Structures

Vladimír Matlach
Palacky University, Olomouc, Czech Republic
vladimir.matlach@upol.cz

In this submission we examine the presence of Menzerath-Altman's Law (MAL) on gene sequences coding proteins with newly designated units of approximately equivalent analogues of those in natural language (Lackova *et al*. 2017 and Faltýnek *et al*. 2019) and its theoretical applications. This law, present in natural languages and associated with its conservation and economy principles, simply states that the larger the construct (a whole),the smaller its elements (or 'components'; see Altmann 1980). In this submission, we show the presence of this linguistic law on the level of DNA coding sequences, precisely on the level of proteins and their secondary structures. We also discuss its potential in practical applications: (1) proteins with a damaged secondary structure (e.g., which cause proteopathic disease) may be so varied in the law that they could easily be identified as outlier values or anomalies; (2) that the non-coding sequence of the proteins would, by predicting their hypothetical secondary structures, be so stray to the law that they would be readily recognizable from real proteins. The first application would lead to the possibility of identifying damaged proteins or more natural annotations of secondary protein structures in the case of several available versions from different prediction software. The second of the intended applications would then be able to detect sequences coding the proteins within the DNA. Verifying that this law is present at the level of proteins and domains leads to theoretical advancement and, above all, to the testing of both applications. For the purpose of testing and maximizing the effectiveness of individual applications, however, the new formal model of the MAL, which surpassed statistically its predecessors, was identified as more advantageous. This submission supplements existing knowledge about coding sequences, proteins, and the existence of MAL in genetics (see e.g. Baixeries et al. 2013, Li 2012 or Shahzad 2015) and its findings, including the hypothetical applications, can be perceived as beneficial.

Bibliography

Altmann, Gabriel. 1980. Prolegomena ToMenzerath's Law. *Glottometrika*, 2, 1–10.
Baixeries, J., Hernandez-Fernandez, A., Forns, N. a Ramon Ferrer-I-Cancho. 2013. The Parameters Of The Menzerath-Altmann Law In Genomes. *Journal Of Quantitative Linguistics*, 20 (2), str. 94-104.
Faltýnek Dan, MatlachVladimír, Ludmila Lacková. 2019. Bases are Not Letters: On the Analogy between the Genetic Code and Natural Language by Sequence Analysis. *Biosemiotics*,1-16.
LackováĽudmila, MatlachVladimír, Faltýnek Dan. 2017. Arbitrariness is not enough: towards a functional approach to the genetic code. *Theory in Biosciences*, 136,3-4, 187-19.
Li, Wentian. 2012. Menzerath's Law At The Gene-Exon Level In The Human Genome. *Complexity* 17 (4), 49-53.
Shahzad, Khuram, Jay E. 2015. Mittenthal a Gustavo Caetano-Anollés. The Organization OfDomains In Proteins Obeys Menzerath-Altmann's Law Of Language. *BMC Systems Biology* 9 (1), 44.

Seventh International Conference in Code Biology – Lužnica (Croatia) 31 August – 4 September 2021

Living robots: how close are we to produce hardware-based artificial life?

Wanderley Dantas dos Santos
State University of Maringa, Department of Biochemistry
Maringa, Parana, Brazil
wdsantos@uem.br

Machines are traditionally classified in simple and composed devices propelled by animal force, self-propelled (automatic) and autonomous machines. Such a classification is convenient reflecting the evolution of machines throughout time. In turn, Maturana and Varela proposed that besides those man made or artificial machines, living beings forms another class of natural machines. Although that living machines are known to evolve along time, even the simplest living beings are autopoietic, i.e. they can build their own components and boundaries. In turn, artificial machines, from any of the classes described above, are allopoietic i.e. they produce something different from themselves. Recently, I have criticized the theory of autopoiesis for explicitly exclude the importance of information for explaining the living phenomenon. Summing up, a piece of information (bit) is a decision between two possibilities. In turn, manufacturing (*poiesis*) is defining the specific position of each part of a product (e.g. the atoms that form a specific organic compound)among all set of other possibilities. Therefore, the specific shape and composition that defines any tridimensional body can be defined as its tridimensional information. The understanding that autopoiesis is inextricably linked to information has implications so profound that I decided to distinguish the concept by the term *semiopoiesis*. The artificial production of a semiopoietic unity using biochemical compounds implies the creation of (wet) artificial life. Here, I discuss how do-it-yourself technologies are getting us close to creation of hardware-based a-life. A fast search in the internet shows projects and STL (stereolithography) files to print and assembly robot arms using standard 3D printers as well as a variety of Arduino compatible robot arms-like 3D printers. It implies that without great innovation breakthroughs it is possible to develop a simple machine capable to produce and assembly its own parts. A basic version should be able to recognize, pick up and assemble the most complex components made available to feed the machine 'self-production'. More advanced technologies would allow to generate smaller machines able to print their own electronic pieces as motors and controllers using simpler materials and different forms of energy. Semiopoietic machines could also be able to reproduce by building a second 3D-printer-assembler. If programmed to cooperate with each other, those machines could generate meta-structures capable to accomplish different kinds of work, no human maintenance needed. They could provide services for homes, cities and unfriendly environments: poisonous mines, deep oceans, space operations and even for exploration or colonization of other planets. From a biological point of view, such a technology will consist in a major transition in evolution of living beings where intelligent life becomes able to produce artificial forms life. In turn, development of semiopoietic hardware would revolutionize the concept of manufacture with a much more dramatic impact in society than the 4.0 industry. The very concept of manufacture would become obsolete since products will be self-built, self-maintained and even self-programmed either evolving by natural selection or projected by AI. All this, right there around the corner.