DNA-Error Control Code May Be Unstructured

H. M. DE OLIVEIRA[a], N.S. SANTOS-MAGALHÃES[b]
[a]Electronics and Systems Department, [b]Biochemistry Department,
Federal University of Pernambuco
C.P. 7.800, 50.711-970, Recife – PE
BRAZIL
hmo@ufpe.br nssm@ufpe.br http://www2.ee.ufpe.br/codec/publicacoes.html

Abstract: - What are introns for? In spite of the fact that they may protect the integrity of the genes, it is almost consent that they have no obvious error control function. This paper answers, to a certain extent, a number of inquiries about introns. It is enlightened why the difference between eukaryotic and prokaryotic transcription arise. Arguments in opposition to the "introns-early theory" are presented. It is suggested that introns have gradually took place in eukaryotic cells as a (Darwinian) probabilistic mechanism to protect replication. We elucidate the mechanisms involved in the augmenting the reliability of the DNA replication by unveiling a nonconforming way to put redundancy to work. Biological evolutionary codes match Shannon's paradigm: they are long truly random codes. On the contrary with current ideas about random codes, they have an uncomplicated decoding algorithm. It is suggested therefore that effort for discovering introns' functional error-control properties should not be made, nor should code structure be searched behind introns.

Key-Words: - introns, biological evolution, error-control codes, theoretical biology.

1 Preliminaries: About Introns

The astonishing reliability by which deoxyribonucleic acid (DNA) has been preserved through ages implies that cell's replication machinery have to ensure against copying mistakes. The replication machine is self-correcting and operates with a mean of 1 error per 10^7 nucleotides copied. Around 99% of such errors are corrected by the DNA mismatch repair mechanism, thereby resulting 1 error per 10^9 nucleotides copied [1,2]. Albeit the entire DNA repair process hardly ever fails, such a repairing can be unsuccessful yielding a change for the next replication round. This is named a mutation. DNA is thereby continually suffering damage in cells. Chemical mutations, spontaneous changes (deamination and depurination) and irradiation account for nearly all alterations [3]. In humans, the mismatch repair failure was recognized as having compelling relation with some species of cancer and aging [2,4]. If, for one hand, it is desirable to allow a diminutive number of mutations in order to prospect the universe of genomes of lower beings, it is as well for other side strongly required to control replication errors in higher organisms. This modus operandi has been playing a crucial role in the theory of evolution [5-7]. Regions of a genome holding more understandable functional properties are the genes. A gene is defined as a DNA segment that is expressed to yield a functional product, which may be RNA or a polypeptide chain. In prokaryotic cells, polypeptide chains are coded in a continuous disposition of 'codons'. During decades, it was thought that genes of higher organisms were also continuous. Nevertheless, eukarya do contain puzzling structural features. In fact, in late 70's, cell biologists found a singular behaviour of the RNA in nucleus from that of bacterial mRNA, and discovered the mosaic of genes in eukaryotic organisms [8].Unexpectedly, their gene organisation is structurally and functionally dissimilar [3,9]. Most eukaryotic genomes contain not only functional genes, but also large amounts of DNA sequences that do not code for proteins. Many noncoding sequences lie between genes (spacer sequences). In prokaryotes, proteins are encoded as a rule by an uninterrupted stretch of DNA sequence. In contrast, most eukaryotic genes have their coding sequences interrupted by noncoding regions (the so-called *introns*, for *intervening nontranscribed sequences*). 'Introns' are usually longer than the 'exons'. They are broadly distributed, some less than 50 bp, other more than 250,000 bp; whereas 'exons' have typical size ranging from 50 to 600 bp (average 300 bp). 'Introns' are more prevalent in higher eukaryotes, atypical in prokaryotic genes, uncommon in lower eukaryotes. Individual genes of eukaryotes are considerably isolated, with long stretches of DNA between one gene and the next. Despite some attempts in understanding the biological role of 'introns' [10], no recognized functions were found. Nevertheless, they account for a substantial fraction of the genome of higher eukaryotes [11]. When the

mRNA produced in eukaryotic cells, the entire genome is first transcript into a long RNA molecule (the primary transcript). Before RNA leaves the nucleus, all of the 'intron' sequences are removed, and the 'exons' joined (a shorter RNA uninterrupted coding sequence is then generated.) This process is the so-called RNA splicing [8]. The unnecessary transcription of spacer-DNA regions, which encode no genetic instructions, is thus avoided taking it away by splicing. The genome is transcribed to produce a sequence from which 'introns' are removed, so merely 'exons' are built-in in the mRNA. The protein coding follows via the genetic code (see [12] for novel representations for the genetic code). The origins and function of 'introns' are often said to remain enigmatic [3]. 'Introns/exons' structures of genes are incredibly ancient, certainly predating the divergence of plants and animals one billion years ago [13,14]. The eukaryotic and prokaryotes cell handled transcription in a different way. How did this difference arise? We try to come back with these inquiries on basis of the evolution theory. The lines presented through this paper support that 'introns' content have gradually augmenting in eukaryotic cells as a (Darwinian) probabilistic mechanism to protect replication. They increase redundancy, protecting as a consequence the genome. Such a redundancy mitigates the damaging effects of mutations on eukaryotic cells. This paper is organized as follows. The current section begins with a brief review on coding and noncoding regions of a genome. The significance of 'introns' and theories about their emerging are then focused. Attempts to find a structure of error-correcting codes are sketched in the next section. Finally, biological error-control coding schemes designed as a product of biological evolution are discussed.

2 The Significance of 'Introns'

What roles play the 'introns'? It is consent that they have no obvious function -- in spite of the fact that they may protect the integrity of the genes. It has also been argued [3] that the hypothesis that "every one of 'introns' are simply selfish DNA" seems untenable since it would otherwise be difficult to rationalize why the evolution of splicing machinery offered any selective advantage over the mere exclusion of the split genes (*sic*). Several kinds of noncoding DNA contribute to the genomic complexity of multicellular eukaryotes. A large amount of the total DNA is composed by nucleotide sequences that are repeated countless times in the genome [15]. Highly repetitive sequences are tandem arrays of thousand of copies of short sequences (5-200 bp). Moderately repetitive sequences (150-300 bp) repeats scattered throughout the genome, and are classified as SINES (short interspersed elements) or LINES (long interspersed elements.) In the humans, they account for 13% and 21% of the genome, respectively [11]. Repetitive DNA has commonly been regarded as "junk-DNA", vestiges of evolutionary sidetracks. Other kind of noncoding DNA is the 'intron', which account roughly for 26% of the human genome. It has been proposed that 'introns' have had some essential function at an earlier stage of genetic evolution that is no longer important, i.e. 'introns' may be genetic fossils. This paper is in opposition to such an idea. It was argued that most of them appear not to make useful contribution to the cell [9,p.145]. 'Modern introns', according to Gilbert's hypothesis [13], are the remnants of a process that facilitated protein evolution. Nevertheless, they must yet play a prominent role; otherwise they gradually had disappeared according to the natural selection. We try to offer here a feasible explanation for their presence in the DNA of higher eukaryotes. A widespread hypothesis is the introns-early theory, which proposed that 'introns' might have been selectively eliminated from the genes of lower organisms. If they are remnants of a primordial gene shuffling process, why are they absent (or near absent) in lower form of life? 'Introns' may be selectively eliminated in these species [3, p.1146]. It was guessed that higher organism has had much less selective pressure for 'intron' elimination just to stay alive this theory. The ideas argued here match another hypothesis, which reports that 'introns' may have arisen in eukaryotic genes. In this paper we elucidate the mechanism involved in the augmenting the reliability of the DNA replication processing via 'introns'. It is commonsense that prokaryotic preceded eukaryotic cells. Viruses and bacteria have a high fecundity and few gene families; have little or almost no need for protection. In contrast, plants and animals have high permanency. Ever since the organisms become much complex, many survivors were, according to the natural selection, those much robust to critical mutations. Our hypothesis is that eukaryotes were then gradually augmenting their amount of 'introns', whose aim is solely to save 'exons' from harm. Why lower organisms such as viruses suffer so many mutations? Their genetic information has little protection (no 'introns') so they may vary from one generation to another relatively close. In contrast, higher organisms are much more sensitive to mutation due to their complexity. A lot of functions are interrelated and a failing in a single functioning may turn out

unfeasible the development of the living system. Many mutations in multicellular systems leaded to defective run and often to death. As a consequence, early complex organism gradually had 'introns' introduced in their genome, since the natural selection kept the most shielded organisms, i.e. those that had a quantity of suitable 'introns'. The average human 'intron' is 35,000 bp long, and there are only ~1.5% protein-coding genes [16]. It can therefore be thought that more complex species should require more 'intron-content'.

3 Looking for the Structure of a Genomic Error-Control Code

All cells have multiple DNA repair systems. Not only known 'proof-reading' mechanisms [2] and post-replicative repair systems [2-3,9], but also strict sense error-correcting codes have been speculated to be concerned with genomic error correction. Battail and other hypothesize that error correcting means are actually implemented in the genome replication process so as to account for the extremely faithful inheritance of information since the very deep past [17-19]. Such codes were probably designed by nature as optimised so as to keep the probability of mutations low enough. 'Introns' are surely involved in the process as they are the foremost responsible for introducing redundancy. DNA should comprise powerful error-control coding to achieve the performances promised by the Shannon theorem [20]. Forsdyke hypothesized that 'introns' could be some kind of check parities associated with the message borne by the 'exons' [21]. However no structured (parity check) codes were found. Further attempts to find out the error-correcting mechanism have also been carried out [21-23]. One of the difficulties to understand the genetic error-coding mechanism should be accounted for the structure of the code and the kind of alphabet it is defined over. This choice is arbitrary, indeed. The most obvious choice for the alphabet is the quaternary alphabet {A, T, G, C} mapped as GF(4). Liebovith et al. assumed the algebraic structure $\mathbb{Z}/4\mathbb{Z}$, the ring of integers modulo 4 [23]. Unfortunately no suitable codes have been found over such alphabet choices. Forsdyke also interpreted a sequence of quaternary symbols as simultaneous bearing two independent binary codes over different alphabets [21]. The first, named {R,Y}, R for purine nucleotides and Y for pyrimidine basis. The second alphabet, {2H,3H}, regarding the number of hydrogen bound corresponding to the complementary base pairing, A=T or C≡G. None of such attempts have been successful. Like it happens in the error-control evolution (from parity checking to turbo codes), this historical development, however, gives the main role in designing structured codes, and we belief this way is badly chosen in the DNA background. Essentially, engineers have investigated standard error-correcting strategies, but the genomic error coding may be rather atypical due to possible insertions and deletions (I/D). A quantity of decoding schemes involving I/D has already been proposed (see [24]), but they were not investigated in genome coding framework. Perhaps the first endeavour to escape the classical paradigm was proposed by Battail who suggested *nearest soft codes* as the means from which the channel coding acts [25]. All discussions till now, however, have been supported by the thought of adding controlled redundancy [26]. This paper presents an alternative issue, opposite to a certain extent: an unconventional way to put redundancy to work, which is discussed in the next sections.

4 A Long Trip Towards a Hands-on Random Coding

The capability of the channel coding was categorically established in the monumental Shannon's work —and the random coding argument was its chief idea [20]. The early paradigm of the coding design was deeply biased by the success of the perfect codes (R. Hamming and M. Golay) in the 50's [27]. The minimum distance and the error correcting ability for the algebraic decoding impelled the status of the coding design. At that time most designed codes were short codes and the achieved transmission rates were far from the channel capacity. Although inappropriate for long codes, the criterion of maximizing the minimum distance between codewords single-minded all error correcting code investigations from 50's to 70's. Families of good codes should asymptotically assure a non-null ratio between the minimum distance and the bocklength. The original Shannon's proof however leaded to the statement that almost all codes are good (even without major focus on the code structure) [20]. In spite of this fact, little accomplishments were achieved in designing codes that approach the channel capacity. The belief that imposing a structure for a code should be incompatible with long good codes was then gradually been consolidated. The folk theorem "all codes are good except those we can think of" was then repeatedly alluded to explain the unsuccessful of coding in 60's and early 70's. If this expectation was true, coding seemed to be relegated to an unfeasible proposal. Thinking and designing long

codes remained as a taboo: so long as some structure was added to a code family, no good codes within this family were found. Now, if the only way to obtain good codes is by means of random coding, their absence of structure should turn the decoding task impractical. For a long while such an incompatibility hovered on the coding theorists. If no structured codes achieve capacity then coding is dead. After the triumph of coding in the deep space channel, the motto was turned to coding is dead except for the deep space channel. The first encouragement came in mid-70's from J. Massey, who criticizes the design of the code as an isolated component of the communication system [28]. He brought attention to the fact that achieved code performance was strongly dependent on the modulation, thereby proposing a joint coding and modulation design. These ideas were deeply developed by G. Ungerboeck who creates the "coded modulation" in early 80's [29,30]. "Coding is dead, long life coding." In spite of the accomplishment of Ungerboeck codes, which practically attain the channel cut-off rate R_0, a more arduous pathway should be tracked. Variants of the main theme were proposed such as any code of which we cannot think is good [31]. R. de Buda stated the first evidence that no incompatibility exists between structure and performance of codes in late 80's [32-33]. H. de Oliveira who proved that there exist structured codes that achieve capacity over the Gaussian channel then improved this argument [34-35]. Both results are concerned with a coded-modulation scheme based on lattice codes. Furthermore, their central proofs do not call for the random coding argument —they are mainly based on the celebrated Minkowski theorem of the geometry of numbers [36]. Once more, proofs are no constructive, but from existence. G. Battail was the first to criticize the use of the conventional minimum distance criterion in the search of long codes [37-39]. He proposed as a criterion the proximity measure with the random code distribution, which was referred as the "pseudo-random criterion." This criticism on the existing code design criteria, *vox clamantis in deserto*, opened a path to escape this deadlock: the design of structured codes that mimic the random code. Turbo codes —invented by C. Berrou and A. Glavieux — appeared in 90's to break the barrier of the cut-off rate, approaching capacity [40-42]. Likewise Ungerboeck's breathtaking work, the telecom community was again somewhat sceptical. Its success was however apparent. Battail quickly caught the reason why the performance of turbo codes was so amazing. He realized that such an outstanding performance could probably be explained by the fact that turbo codes mimic the random code [42-43]. This was a decisive step towards the statement: coding abides. In several papers, Battail went backward to the reasoning that "we can think of good codes and even decode them." The rationale was always to mimic the random coding using a deterministic structure. The channel coding for sustaining the life seems associated with Shannon's groundwork: DNA may act as a random code.

5 Genomic Error-Control Codes as a Product of Biological Evolution

In many cases, terms like "error-control coding" should be preferred to "error-correcting coding", although the latter is widespread. Trying to bear out the error control ability implicated in the DNA processing, it has been proposed that the constraints imposed by the DNA structure could act as some sort of coding process [18,21-22]. This paper is in opposition to this reasoning. Standard error correcting codes are usually designed by imposing constraints on the sequences. However, we think that Darwinian mechanisms for protecting DNA information may be quite different. That is probably why no one had success in unveiling the coding structure of DNA sequences. No parity rules should be looked for. No block or convolutional structure over groups or rings should be searched [26]. Why using structured codes? The standard answer is based on the (mislead) belief that the decoding of random code is unfeasible. At the receiver end, each word of the codebook has to be compared with the received word. Due to the lack of structure, an exhaustive search would be supposed to be required. The proposal of providing codes with a structure is to make decoding undemanding enough. This has been the major trend in the search of codes, and explains that random coding remained merely a theoretical tool for proving the channel capacity theorem [20,44]. This opinion has long been shared by virtually every one of coding theorists. Battail proposed the formerly escape from this paradigm [45-46]: Deterministic means to mimic random coding, the so-called random-like deterministic codes. Biologists have long been made unsuccessful efforts to discover functional properties for 'introns'. For their side, coding theorists have in vain tried to make out some structured code behind 'introns'. They were both pursuing adverse traces by keeping themselves deeply rooted to their customary paradigms. The need for finding usefulness for noncoding regions of genome revealed the thinking

of biologists who were essentially looking for chemical or biological functionality. However, we believe that the secret of 'introns' lies on the natural selection: 'introns' were the spontaneous mechanism of introducing uncertainty [47]. We are convinced that the following striking analogy can be valuable. In a battle, a crucial payload is to be sent to the front. If the only way is sending it through the battlefield, it should not be directly dispatched. Many fake-cargos could be added, and the relevant one will be hidden among them. If the enemy (noise, mutation) hardly tries to intercept this crucial delivery, he can now probably not succeed due to the amount of uncertainty added to the process. Many ineffective cargos (junk-cargos or 'introns') will be hit, but the main one will probably be missed. This is certainly the means adopted by the genome. This strategy is currently used in the safeguard of authorities such as Presidents of some nations (to include uncertain routes and second self.) To sum up, our hypothesis is that the DNA coding is essentially the truly random code. This can be supported by the small dissimilarity between their *spectrograms* or *scalograms* [48]. Additionally the *a^2grams* and *codongrams* of genomes also exhibited patterns resembling to those of a random sequence [49]. The reasoning that strongly structured codes must be used to guarantee an undemanding decoding algorithm lacks of foundation: DNA coding has trivial decoding scheme based on the asynchronous start-stop protocol. DNA code meet Battail's close-to-random criterion and its "decoding" is easily implemented. As a consequence, we quote Battail [17]: "Nature appears as an outstanding engineer…"

6 Closing Remarks

The community is at large not especially concerned with the present focus, although information theory approach could be valuable in many Biochemical problems [24,50]. This is partially a consequence of the deep multidisciplinary nature of the theme, which may turn a bit hard to grasp the contributions. Nevertheless, it is point in time to make people mindful to think about it. A multidisciplinary endeavour is required to progress on the arguments presented in this paper. After reading this paper, a coding theorist could ask: "Well, where is the code?" If he feels a bit disappointed, it suffices to celebrate: *Shannon was right*… Finally, we feel a bit striking that the elucidation of the protecting mechanisms of genetic information as well as the vindication about how 'introns' appeared can only be made on the basis of the evolution theory and natural selection. … Under a similar title of an early work [37], we claim: "Natural world can think of practical random codes and even decode them." Simulations to scrutinize the performance of these schemes over the additive Gaussian noise channel should be carried out to corroborate the chief ideas of this paper and to provide further understanding about DNA-like error-control schemes.

Acknowledgments:-The authors thank Prof. F.M. Campello de Souza for insights into the actual relevance of the probabilistic mechanism in Nature. HMdO also is grateful to Prof. G. Battail for such a deep influence in his way of appreciating Science.

References:

[1] B. Alberts, D. Bray, A. Johnson, J. Lewis et al., *Essential Cell Biology*. Garland Pub., 1998.

[2] D.L. Nelson, M.M. Cox, *Lehninger Principles of Biochemistry*. 3rded. Worth Publishers, 2000.

[3] D. Voet, J. Voet, *Biochemistry*, 2nded., Wiley, 1995.

[4] P. Modrich, Mismatch Repair, Genetic Stability, and Cancer, *Science*, 266, pp.1959-1960, 1994.

[5] J.E. Darnell, W.F. Doolitle, Speculations on the Early Course of Evolution, *Proc. Natl. Acad, Sci*., vol.83, pp.1271-1275, 1986.

[6] L.E. Orgel, The Origin of Life – a review of facts and speculations. *Trends Biochem. Sci*., vol. 23, pp.491-495, 1998.

[7] M. Ridley, *Evolution*, Midsomer Norton: Oxford Univ. press, 1997.

[8] P.A. Sharp, On the Origin of RNA Splicing and introns, *Cell*, vol.42, pp.397-400, 1985.

[9] G.F. Cooper, *The Cell: A Molecular Approach*, 2nd ed., ASM press, Sunderland, MA, 2000.

[10] W.W. Gibbs, The Unseen Genome: Gems among the Junk, *Sci. Am.*, 289, pp.46-53, 2003.

[11] C. Dennis, R. Gallagher, *The Human Genome*, Nature-Palgrave, Hampshire, 2001.

[12] H.M. de Oliveira, N.S. Santos-Magalhães, The Genetic Code Revisited, *Lecture Notes in Computer Science*, N° 3124, Springer Verlag, pp. 526-531, 2004.

[13] G. Gilbert, M. Marchionni, G. McKnight, On the Antiquity of Introns, *Cell*, vol.46, pp.151-154, 1986.

[14] M. Marchionni, G. Gilbert, The Triosephosphate Isomerase Gene from Maize: introns antedate the palntanimal divergence, *Cell*, vol.46, pp.133-141, 1986.

[15] R.J. Britten, D. Kohne, Repeated Sequences in DNA, *Science*, vol. 161, pp.529-540, 1968.

[16] N.S. Santos Magalhães, H.M. de Oliveira, Of Protein Size and Genomes, *WSEAS Trans. On Biol. And Biomed*., vol.3,Feb., pp.133-138, 2006.

[17] G. Battail, An Engineer's View on Genetic Information and Biology Evolution,*Workshop on Info. Proc. in Cell and Tissues*, Lausanne, 2003.

[18] J. Rzeszowska-Wolny, Is Genetic Code Error-correcting? *J. Theor. Biol*., vol.104, pp.701-702, 1983.

[19] E. May, M.A. Vouk, D.L. Bitzer, D.I. Rosnick, Coding Theory Based Models for Protein Translation Initiation in Prokaryotic Organisms, IPCAT'03, 5th *Int. Workshop on Info. Proc. in Cell and Tissues*, Lausanne, pp.371-389, 2003.

[20] N.J.A. Sloane, A.D Wyner, *Claude Elwood Shannon, Collected Papers*, IEEE Press, 1993.

[21] D.R. Forsdyke, Are Introns in-series Error-detecting Sequences? *J. Theor. Biol*., vol.93, pp. 861-866, 1981.

[22] D.A. Mac Dónail, A Parity Code Interpretation of Nucleotide Alphabet Composition, *Chem. Communic*., vol.18, pp. 2062-2063, 2002.

[23] L.S. Liebovithch, Y. Tao, A.T. Todorov, L. Levine, Is there an Error Correcting Code in the Base Sequence in DNA? *Biophysical J.*, vol.71, pp. 1539-1544, 1996.

[24] L.J. Schulman, D. Zukerman, *IEEE Trans. Info. Theory*, pp.2552-2557, 1999.

[25] G. Battail, Is Biological Evolution Relevant to Information Theory and Coding? *Proc. Int. Symp. on Coding Theory and Applications*, ISCTA 2001, Ambleside UK,pp.343-351, 2001.

[26] S.B. Wicker, *Error Control Systems for Digital Commun. and Storage*, Prentice Hall, NJ, 1995.

[27] T.M. Thompson, *From Error-Correcting Codes though Sphere Packing to Simple Groups*, The Math. Assoc. of America, 1983.

[28] J.L. Massey, Coding and Modulation in Digital Communications, Proc. *Int. Zurich Seminar on Digital Commun.*, Zurich, pp.E2(1)-(4) 1974.

[29] G. Ungerboeck, I. Csajka, On improving data line performance by increasing the channel alphabet and introducing sequence coding, IEEE *Int. Symp. on Info. Theory*, Ronneby, Sweden, p.53, 1976.

[30] G. Ungerboeck, Channel Coding with Multilevel/phase Signals, *IEEE Trans. on Info. Theory*, IT 28, pp.55-67, 1982.

[31] J.T. Coffey and R.M. Goodman, Any Code of which we Cannot Think is Good, *IEEE Trans. Info. Theory*, IT 36, pp.1453-1461, 1990.

[32] R. de Buda, The Upper Error Bound of a new Near-Optimum Code, *IEEE Trans. on Info. Theory*, IT 21, pp.441-445, 1975.

[33] R. de Buda, Some Optimal Codes have Structure, *IEEE J. Select. Areas on Commun.*, SAC7, pp.893-899, 1989.

[34] H.M. de Oliveira, G. Battail, A Capacity Theorem for Lattice Codes on Gaussian Channels, *proc. IEEE/SBT Int. Telecomm. Symp*., pp.1.2.1-1.2.5, Rio de Janeiro, Brazil, 1990.

[35] H.M. de Oliveira, G. Battail, On Lattice Coding for the Gaussian Channel, *Europ.Trans. on Telecommun.*, vol.4,nº2, pp.133-140, 1993.

[36] J.W.S. Cassels, *An Introduction to the Geometry of Numbers*, Springer-Verlag, 1959.

[37] G. Battail, We can Think of Good Codes, and even Decode them, Eurocode 92, *CISM lectures*, nº339, Springer Verlag, pp.353-358, 1993.

[38] G. Battail, Is Minimal Distance a Good Criterion? *Lecture Notes in Computer Sciences*, nº 338, Springer Verlag, pp.325-327, 1989.

[39] G. Battail, Construction explicite de bons codes longs, *Ann. Telecomm.*, vol.44,pp.392-404, 1989.

[40] G. Berrou, A. Glavieux, P. Thitimajshima, Near Shannon Limit, error-correcting Coding and Decoding: Turbo Codes, *IEEE Int. Conf. on Commun.*, ICC'93, vol2/3,pp.1064-1071, 1993.

[41] G. Battail, C. Berrou, A. Glavieux, La capacité d'um canal peut être approchée par des moyens simples,*GRETSI*,Juan-les-Pins,pp.623-626. 1993.

[42] G. Battail, C. Berrou, A. Glavieux, Pseudo-random Recursive Convolutional Coding for Near-capacity Performance, Proc. *GLOBECOM*, vol.4, pp.23-27, Houston, 1993.

[43] G. Battail, Turbo-codes as Random-like Codes, *Mediterranean Workshop on Coding and Info. Integrity*, Palma de Mallorca, Spain, 1996.

[44] H.M. de Oliveira, G. Battail, The Random-coded Modulation: Performance and Euclidean Distance Spectrum Evaluation, *Ann.Telecomm.*, vol.47 nº3-4, pp. 107-124, 1992.

[45] G. Battail, On Random-like Codes, *Lecture Notes in Computer Science*, nº1133, pp.76-94, Springer Verlag, 1996.

[46] G. Battail, Randomness and Structure in Channel Coding, in *Communication and Coding*, pp.30-47, RSP & John Wiley, Jan. 1998.

[47] F.M. Campello de Souza, *Decisões Racionais em Situações de Incerteza*, (Portuguese) UFPE-Editora Universitária, Recife, 2002.

[48] A.A. Tsonis, P. Kumar, J.B. Elsner, P.A. Tsonis: Wavelet Analysis of DNA Sequences. *Phys. Rev.* E 53, pp.1828-1834, 1996.

[49] E.A. Bouton, H.M. de Oliveira, R.M. Campello de Souza, N.S. Santos-Magalhães, Genomic Imaging Based on Codongrams and a^2grams, *WSEAS Trans. on Biol. and Biomed.*, vol.1, pp.255-260, April 2004.

[50] G. Battail, Does Information Theory Explain Biological Evolution? *Europhysics Lett.*, vol.40, pp.343-348, 1997.