Hindawi Publishing Corporation
Abstract and Applied Analysis
Volume 2012, Article ID 439089, 13 pages
doi:10.1155/2012/439089

Research Article

Shannon Information and Power Law Analysis of the Chromosome Code

J. A. Tenreiro Machado

*Department of Electrical Engineering, Institute of Engineering of Polytechnic of Porto,
Rua Dr. António Bernardino de Almeida, 431, 4200-072 Porto, Portugal*

Correspondence should be addressed to J. A. Tenreiro Machado, jtm@isep.ipp.pt

Received 8 June 2012; Revised 17 August 2012; Accepted 21 August 2012

Academic Editor: Dumitru Băleanu

This paper studies the information content of the chromosomes of twenty-three species. Several statistics considering different number of bases for alphabet character encoding are derived. Based on the resulting histograms, word delimiters and character relative frequencies are identified. The knowledge of this data allows moving along each chromosome while evaluating the flow of characters and words. The resulting flux of information is captured by means of Shannon entropy. The results are explored in the perspective of power law relationships allowing a quantitative evaluation of the DNA of the species.

1. Introduction

During the last years the genome sequencing project produced a large volume of data that is presently available for computational processing [1–14]. Researchers have been tackling the information content of the deoxyribonucleic acid (DNA), but interesting questions remain still open [15–21].

This paper addresses the information flow along each DNA strand. For this purpose several statistics are developed, and the relative frequencies of distinct types of symbol associations are evaluated. The concepts of character, word, word delimiter, and phrase are defined, and the information content of each chromosome message is quantified. Power law (PL) relationships emerge in the information locus. PL distributions, often known as heavy tail distributions, Pareto laws, Zipf laws, or others, have been largely reported in the modeling of distinct real phenomena [22–31]. It was recognized [11, 32–34] that DNA has an information structure that reveals long range behavior, somehow in the line of thought of systems with dynamics described by the tools of Fractional Calculus (FC) [35–37]. It is

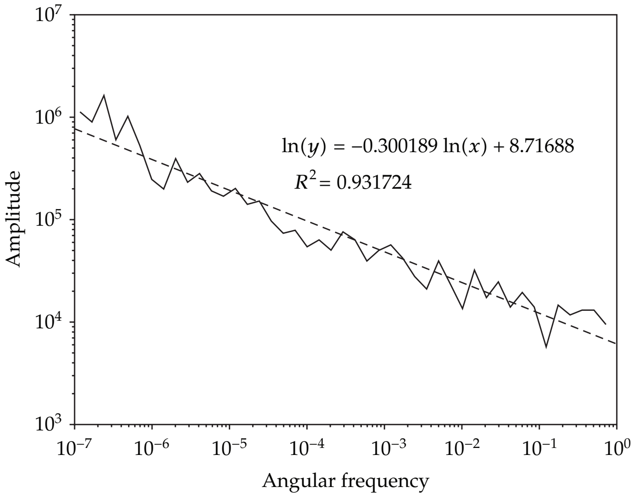


Figure 1: Amplitude of the Fourier transform versus frequency ω for chromosome 1 of the human being (solid line) and PL approximation (dashed line).

known the existence of a strong relationship between FC and PL; nevertheless, up to the present state of knowledge, no formal demonstration supported that observation based on empirical and experimental measurements. Therefore, it is not a surprise that both FC and PL descriptions emerge when analyzing DNA with distinct mathematical tools. In the present study PL descriptions are applied for condensing the charts characterizing the chromosomes of twenty-three species.

Having these ideas in mind, this paper is organized as follows. Section 2 presents the DNA sequence decoding concepts, the mathematical tools and formulates the algorithm that computes the information for each chromosome and species. Section 3 analyzes the DNA information dynamical content of 463 chromosomes corresponding to a set of twenty-three species. Finally, Section 4 outlines the main conclusions.

2. Preliminary Notes on the DNA Information

In the DNA double helix there are four distinct nitrogenous bases, namely, thymine, cytosine, adenine, and guanine, denoted by the symbols $\{T, C, A, G\}$. Each type of base on one strand connects with only one type of base on the other strand, forming the base pairing $A - T$ and $G - C$. Besides the four symbols $\{T, C, A, G\}$, the available chromosome data includes a fifth symbol "N" which is believed to have no practical meaning for the DNA decoding.

For processing the DNA information a possible technique is to convert the symbols into a numerical value. In previous papers was adopted the direct symbol translation $= 1 + i0$, $C = -1+i0, T = 0+i, G = 0-i, N = 0+i0$, where $i = \sqrt{-1}$. We can move along the DNA strip, one symbol (base) at a time. The resulting values form a "signal" $x(t)$ where "t" can be interpreted as a pseudotime. The signal can be treated by the Fourier transform $F\{x(t)\} = \int_{-\infty}^{+\infty} x(t)e^{-i\omega}dt$, where ω represents the angular frequency.

Figure 1 shows one example with the amplitude of the Fourier transform for chromosome 1 of the human being. The frequency interval $10^{-7} \leq \omega \leq 10^0$ is adopted and a PL approximation is superimposed revealing a strong correlation.

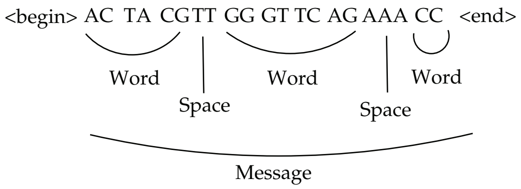


Figure 2: Example of a message when considering $n = 2$, {TT, AA} ≡ "spaces", {TC, TA, TG, CT, CC, CA, CG, AT, AC, AG, GT, GC, GA, GG} ≡ "word characters." Multiple consecutive spaces are considered as a single space.

This technique has, however, one drawback which is the initial assignment of numerical values to the DNA symbols. Therefore, it is important to design an alternative method of analysis avoiding that problem, but, on the other hand, capable of revealing fractional order phenomena. Bearing this strategy in mind, in this paper is adopted an approach based on the histograms of symbol alignment, information theory, and PL approximations.

This study focuses over twenty-three species yielding a space of 463 chromosomes. Therefore, denoting by N_j the number of chromosomes of species $j = 1, \ldots, 23$, we consider the $\{\text{Species}, \text{Tag}, N_j\}_j$ given by {Mosquito (*Anopheles gambiae*), Ag, 5}$_1$, {Honeybee, (*Apis mellifera*), Am, 16}$_2$, {*Caenorhabditis briggsae*, Cb, 6}$_3$, {*Caenorhabditis elegans*, Ce, 6}$_4$, {Chimpanzee, Ch, 25}$_5$, {Dog, Dg, 39}$_6$, {*Drosophila simulans*, Ds, 6}$_7$, {*Drosophila yakuba*, Dy, 10}$_8$, {Horse, Eq, 32}$_9$, {Chicken, Ga, 31}$_{10}$, {Human, Ho, 24}$_{11}$, {Medaka, Me, 24}$_{12}$, {Mouse, Mm, 21}$_{13}$, {Opossum, Op, 9}$_{14}$, {Orangutan, Or, 24}$_{15}$, {Cow, Ox, 30}$_{16}$, {Pig, Po, 19}$_{17}$, {Rat, Rn, 21}$_{18}$, {Yeast (*Saccharomyces cerevisiae*), Sc, 16}$_{19}$, {Stickleback, St, 21}$_{20}$, {Zebra Finch, Tg, 32}$_{21}$, {Tetraodon, Tn, 21}$_{22}$ and {Zebrafish, Zf, 25}$_{23}$.

The DNA information decoding is addressed in this paper, and we start by defining the underlying concepts. The fundamental unit is the "symbol" that, in our case, consists in one of the four possibilities $\{T, C, A, G\}$, while "N" is simply disregarded. Each "character" is represented by an n-tuple association ($n = 1, 2, \ldots$) of the 4 symbols, resulting in a total of 4^n possible symbols per character. For example, with $n = 2$ we get a maximum of 4^2 characters represented by the 16 two-symbol sequences {TT, TC, TA, TG, CT, CC, CA, CG, AT, AC, AA, AG, GT, GC, GA, GG}. The sequences are obtained when moving sequentially along the DNA. The characters may have different significance and are divided into two classes, namely, characters with relevant information, to be denoted in the sequel as "word characters," and delimiters denoted as "spaces." Therefore, joining consecutive "word characters" yields a "word," that ends in the presence of one or more consecutive "spaces" (i.e., multiple spaces are considered as a single space). When the complete association of consecutive words is fulfilled, we obtain a "message."

Figure 2 depicts a simple example of a message with 21 symbols and 3 words. The message {ACTACGTTGGGTTCAGAAACC} is processed according to the proposed scheme for $n = 2$ and considering the 2 sequences {TT, AA} as spaces, and the 14 sequences {TC, TA, TG, CT, CC, CA, CG, AT, AC, AG, GT, GC, GA, GG} as characters. Therefore, the resulting words are {AC TA CG}, {GG GT TC AG} and {CC}.

We verify that we may have words with different lengths and that it is considered as a single space any repetition of spaces. The message finishes when the end of the DNA strand is attained, and, therefore, it is not considered the case of multiple messages for each chromosome.

After defining the concepts for symbol, character (with the categories of word character and space), and message, we need to establish the numerical value to be adopted by n and the method for measuring the information. In what concerns n no *a priori* optimal value is considered. Therefore, in the experiments is analyzed the influence when going from $n = 1$ up to $n = 12$, or, correspondingly, when going from 4^1 up to 4^{12} symbols per character. This evaluation is performed for one chromosome. Based on this first assessment, given the huge computational load required by high values of n, the set of twenty-three species, totalizing 463 chromosomes, is analyzed for $n = \{1, \ldots, 8\}$. In what concerns the information measurement it is adopted the Shannon information [38–49] $I_i = -\ln(p_i)$ where I_i represents the quantity of information of event i that has a probability p_i. In this topic we can refer to [50] calculating also the Shannon information for short DNA words of differing lengths, where the authors find that genomes share universal statistical properties. It is also worth mentioning that other entropies, such as the Rényi, Tsallis, and Ubriaco definitions [51, 52] were tested. Nevertheless, experiments with these expressions and distinct numerical values of the parameters did not reveal any significant conceptual difference. Therefore, for simplicity in the sequel it is adopted merely the Shannon definition.

In our case, for a n-tuple symbol encoding, the occurrence of the ith character within the 4^n set has probability ${p_i}^{\text{char},n}$ leading to information $-\ln({p_i}^{\text{char},n})$, and, therefore, the total information content of a word $I^{\text{word},n}$ yields

$$I^{\text{word},n} = -\sum_{i=1}^{m} \ln\left({p_i}^{\text{char},n}\right), \tag{2.1}$$

where m represents the total number of word characters including the first space. In fact, it was numerically evaluated the effect of including, or not, the space information but, due to its low importance, the final effect is negligible. Therefore, it is considered the inclusion of one space as the information for delimiting the word, while further consecutive repetitions of spaces are disregarded.

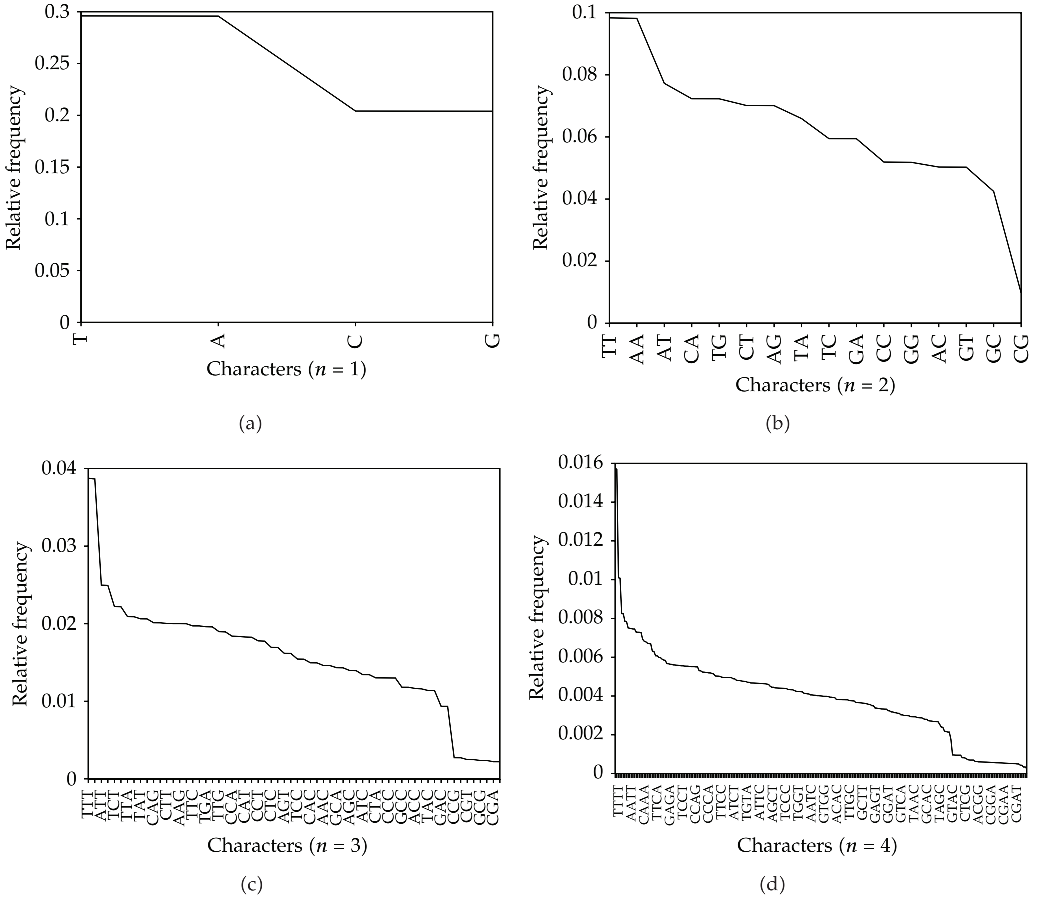
The message information is the sum of all word information:

$$I^{\text{mes},n} = \sum_{i=1}^{r} {I_i}^{\text{word},n}, \tag{2.2}$$

where r denotes the total number of words included in the message (i.e., the chromosome).

The information measurement requires the knowledge of ${p_i}^{\text{char},n}$. While we can expect an equilibrium of probabilities for $n = 1$, that may be not true for larger values of n. Therefore, in the sequel it is adopted a numerical procedure that starts by reading the chromosome message based on the n-tuple character setup leading to the construction of one histogram per chromosome. In the set of 4^n bins are chosen, by inspection, those that are more frequent (and have smaller information content) for the role of spaces. In a second phase, the relative frequencies, which are adopted as approximants to the probabilities, and the information values (2.1) and (2.2) are calculated numerically while traveling along the DNA strand.

This strategy does not consider some *a priori* optimal value of n. Therefore, as mentioned previously, several distinct values of n will be studied before establishing any conclusions.



(a)

(b)

(c)

(d)

Figure 3: Histograms for Ho12 and $n = \{1, 2, 3, 4\}$.

3. Capturing the DNA Information

We start by considering Human chromosome 12 (Ho12) and $n = \{1, \ldots, 12\}$. This chromosome is represented by a medium size file (130 Mbytes) and may be considered a good compromise between length and computational load.

Figure 3 depicts the histograms for $n = \{1, 2, 3, 4\}$ where, for simplifying the visualization, the characters are ordered by decreasing magnitude of relative frequency. For the histograms construction two counting methods were envisaged: (i) counting with disjoint set of n symbols and (ii) counting the sets while sliding one symbol at a time. At first sight it seems that (i) is the most straightforward, but if we consider that we do not have reliable information for starting and synchronizing the counting, then method (ii) is more robust and, therefore, is adopted in the sequel.

Figure 4 shows the word information dynamics when travelling along the Ho12 strand for $n = \{1, 2, 3, 4\}$. We observe the existence of quantum information levels that somehow vanish when n increases. This is due to finite number of quantifying levels of information that occur before a space terminates a word. The number of quantum levels increases with n while

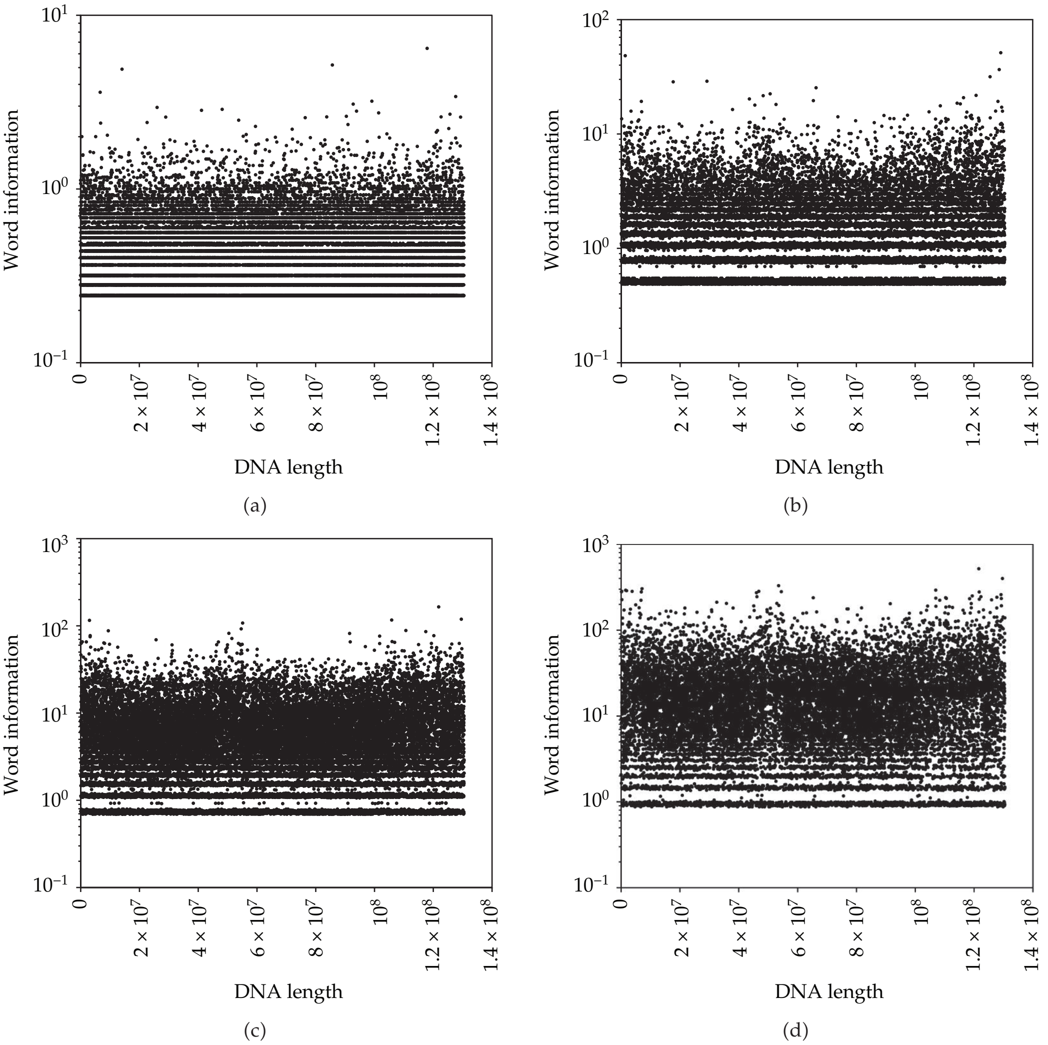


Figure 4: Word information *versus* length for the Ho12 and $n = \{1, 2, 3, 4\}$.

the length of each word increases. Besides this interesting effect, we also note a considerable randomness and a uniform behavior along all length of the strand.

The total chromosome information, the number of words N_w, and the average word information I_{av} versus n are depicted in Figures 5(a) and 5(b). We verify a maximum of the total chromosome information for $n = 3$. For larger values of n the information decreases slightly due to the effect of dropping out repeated consecutive spaces. Therefore, we can say that large values of n seem to lead to a slightly better estimate of the total information content, while the cases of $n = 1$ or $n = 2$ lead to an inferior measurement process. We also observe that the number of words decreases with n but its average information varies in the opposite way. Therefore, it is relevant to plot one variable against the other, with n as parameter (Figure 5(c)). A PL trendline approximation demonstrates that the two quantities are inversely proportional. In fact, we get numerically $I_{\text{av}} = a{N_w}^b$ with $a = 2.07\,10^8$,

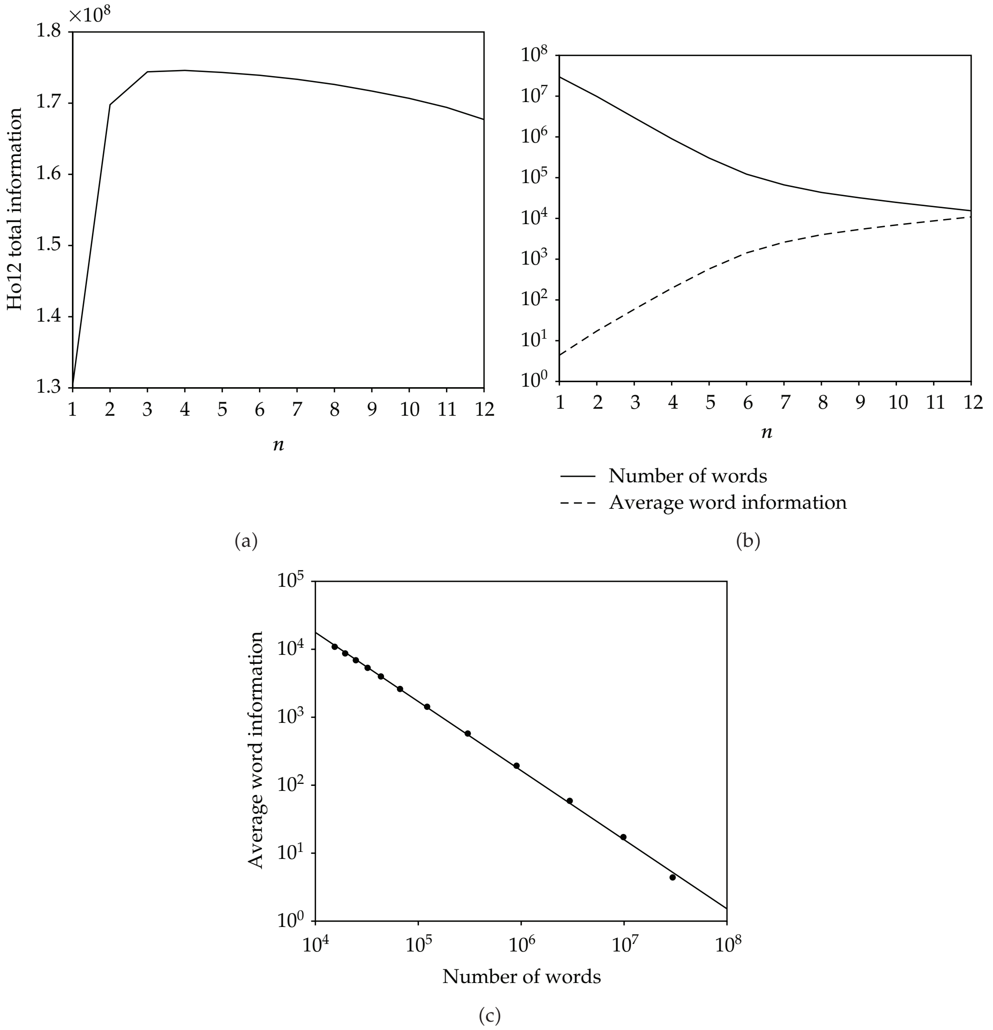


Figure 5: Chromosome Ho12: (a) total information *versus* n, (b) average word information and number of words *versus* n, (c) average word information *versus* number of words.

$b = -1.02$. For the rest of the chromosomes it was observed a similar type of behavior, but with different numerical values for the parameters.

For other values of n the resulting histograms reveal identical characteristics, namely, two characters with a very large relative frequency (depicted at the left part of the histograms of Figure 3). Furthermore, experiments with other chromosomes lead to similar results. The two characters are simply a succession of symbols A or T and the corresponding n-tuples (i.e., $A \cdots A$ and $T \cdots T$) are adopted in the sequel as "spaces."

Figure 6 shows the total information, that is, the information resulting from summing the information of all the chromosomes of each species *versus* the corresponding number of chromosomes, for character encoding with $n = 8$. We observe a weak correlation between both variables.

Figure 7 shows the length of each chromosome L_i^{crom} versus its information content $I_i^{\text{crom},n}$, $i = 1, \ldots, 463$, estimated by the proposed method with $n = 8$. In this case we observe

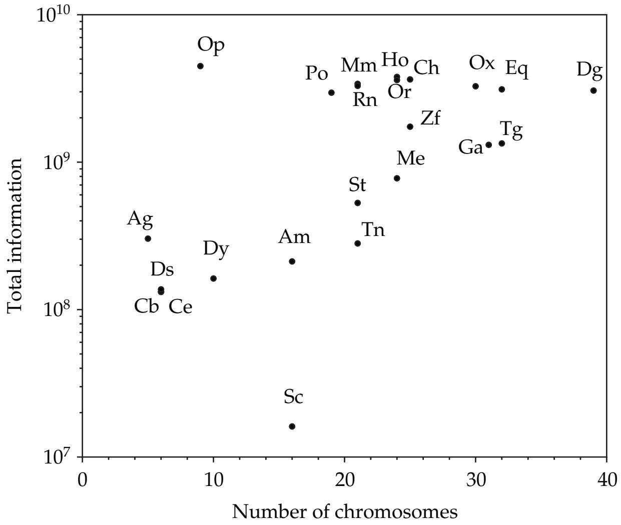


Figure 6: Total information for each species versus the number of chromosomes with $n = 8$.

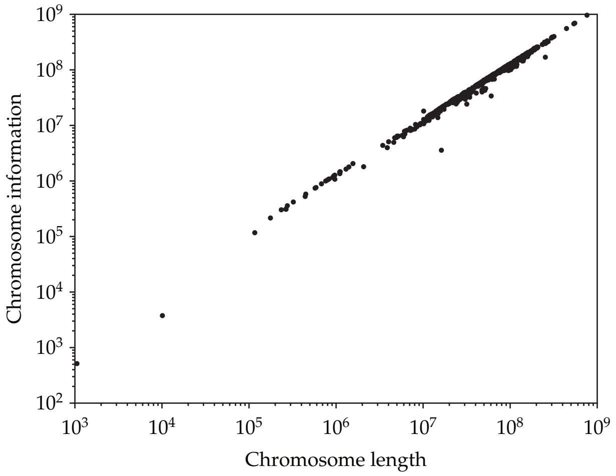


Figure 7: Chromosome length L_i^{crom} versus its information content $I_i^{\text{crom},8}$, $i = 1,\ldots,463$, $n = 8$.

a strong correlation between both variables, meaning that the implementation of the DNA code has a large similarity between all species. In fact, we can calculate a PL trendline over the 463 chromosomes yielding the relationship $I_i^{\text{crom},8} = 0.79(L_i^{\text{crom}})^{1.03}$.

Bearing these ideas in mind it was decided to explore the PL behavior, that is, the relation $I_{\text{av}} = aN_w{}^b$, $a > 0, b < 0$, of the average word information I_{av} versus the number of words N_w (with n as parameter) per chromosome. The extensive evaluation of the 463 chromosomes for $n = \{1,\ldots,8\}$ leads to the locus (a, b) of the PL trendline depicted in Figure 8. The point for chromosome DyYh is not included to allow a better visualization of the remaining set of points. Moreover, the individual chromosome labels are not included to make the plot more readable.

We verify that the map produces clear patterns, not only by grouping the chromosomes of each species but also by the relative positioning of the different species.

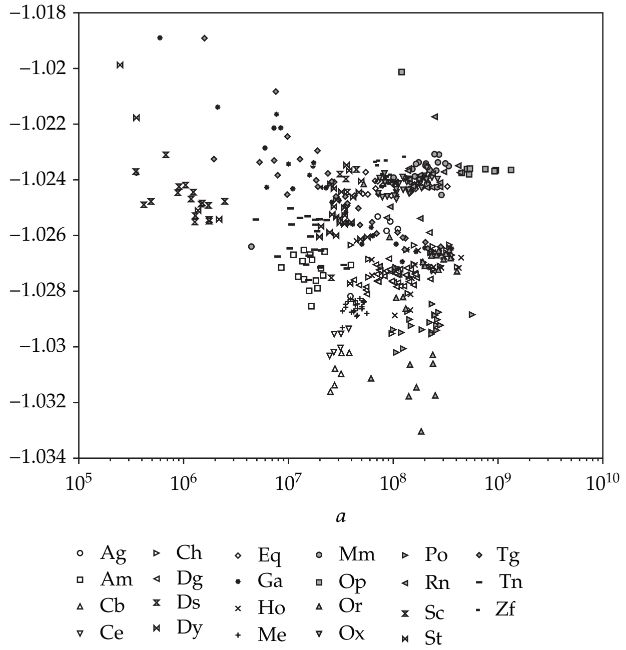


Figure 8: Locus (a, b) of the power law parameters for the 463 chromosomes and $n = \{1, \ldots, 8\}$.

Nevertheless, the large number of points complicates the visualization. Therefore, it was decided to represent each species by a single point having for coordinates the geometric and arithmetic averages of parameters a and b, respectively. Figure 9 depicts the resulting locus where is now easier to analyze the previously mentioned relations. The microchromosomes Ga32 and Tg16, which have a very small base pair counting, were not included in the calculations because they significantly disturb the results.

We verify the emergence of clusters that are in reasonable accordance with phylogenetics, going from the less "complex" species at left up to the most "complex" species at the right. The cluster of mammals is at the right and includes the subcluster of primates {Ho, Ch, Or}, with Ch closer to Hu than Or. In the rest of mammals it is interesting to see Po close to the primates and the position of the marsupial Op relatively distant from the placental mammals. In what concerns the rest of the points we notice Cb close to Ce and, in a middle position, the clusters of birds {Ga, Tg}, fishes {Tn, St, Me, Zf}, and insects {Dy, Ds, Am, Ag}.

In conclusion, the proposed information measure leads to an assertive and quantitative classification of chromosomes and species. Furthermore, it can be further explored for decoding in more detail other aspects of the DNA code in association with the FC tools.

4. Conclusions

Chromosomes have a code based on a four-symbol alphabet, and it can be analyzed with methods usually adopted in information processing. The information structure has resemblances to those occurring in systems characterized by fractional dynamics. Nevertheless, schemes based on assigning numerical values to the DNA symbols may deform the information, and alternative methods that avoid such problem need to be implemented. In this paper it was proposed a scheme based on the Shannon information theory. Bearing these ideas in

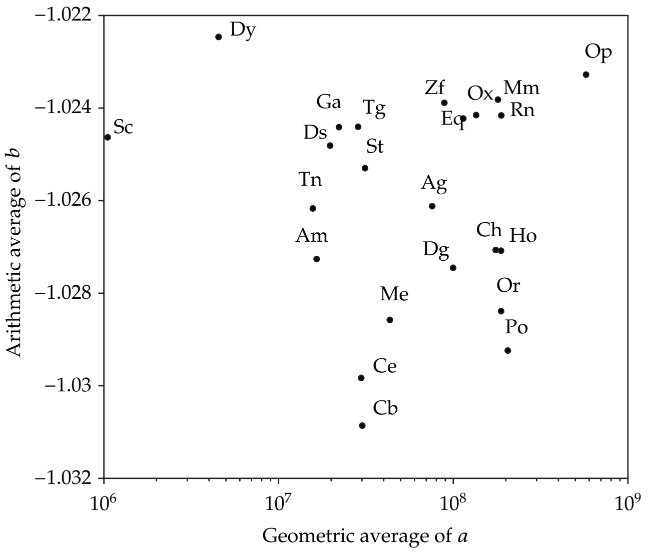


Figure 9: Locus of geometric average of a versus the arithmetic average of b for the twenty-three species.

mind, the chromosomes were processed in the perspective of a PL relationship between the average information and the total number of words, for distinct values of character encoding. For condensing the information an averaging of the PL parameters was also adopted. The resulting locus revealed the emergence of clearly interpretable patterns in accordance with current knowledge in phylogenetics. The proposed methodology opens new directions of research for DNA information processing and supports the recent discoveries that fractional phenomena are present in this biological structure.

Acknowledgments

The authors thank the following organizations for allowing access to genome data: Gambiae Mosquito (The International Anopheles Genome Project), Honeybee (The Baylor College of Medicine Human Genome Sequencing Center, http://www.hgsc.bcm.tmc.edu/projects/honeybee/), Briggsae nematode (Genome Sequencing Center at Washington University in St. Louis School of Medicine), Elegans nematode (Wormbase, http://www.wormbase.org/), Common Chimpanzee (Chimpanzee Genome Sequencing Consortium), Dog (http://www.broad.mit.edu/mammals/dog/, Lindblad-Toh K., et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005 Dec 8; 438 : 803-19), *Drosophila simulans* (http://genome.wustl.edu/genomes/view/drosophila_simulans_white_501), *Drosophila yakuba* (http://genome.wustl.edu/genomes/view/drosophila_yakuba), Horse (http://www.broad.mit.edu/mammals/horse/), Chicken (International Chicken Genome Sequencing Consortium Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 2004 Dec 9; 432(7018): 695-716. PMID: 15592404), Human (Genome Reference Consortium, http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/), Medaka (http://dolphin.lab.nig.ac.jp/medaka/), Mouse (Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature, 420, 520–562 (2002),

http://www.hgsc.bcm.tmc.edu/projects/mouse/), Opossum (The Broad Institute, http://www.broad.mit.edu/mammals/opossum/), Orangutan (Genome Sequencing Center at WUSTL, http://genome.wustl.edu/genome.cgiGENOME=Pongo%20abelii), Cow (The Baylor College of Medicine Human Genome Sequencing Center, http://www.hgsc.bcm.tmc.edu/projects/bovine/), Pig (The Swine Genome Sequencing Consortium, http://pig-genome.org/), Rat (The Baylor College of Medicine Human Genome Sequencing Center, http://www.hgsc.bcm.tmc.edu/projects/rat/, Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428(6982), 493–521 (2004)), Yeast (Saccahromyces Genome Database, http://www.yeastgenome.org/), Stickleback (http://www.broadinstitute.org/scientific-community/science/projects/mammals-models/vertebrates-invertebrates/stickleback/stickleba), Zebra Finch (Genome Sequencing Center at Washington University St. Louis School of Medicine), Tetraodon (Genoscope, http://www.genoscope.cns.fr/), and Zebrafish (The Wellcome Trust Sanger Institute, http://www.sanger.ac.uk/Projects/D_rerio/).

References

[1] R. T. Schuh and A. V. Z. Brower, *Biological Systematics: Principles and Applications*, Cornell University Press, 2nd edition, 2009.

[2] H. Seitz, *Analytics of Protein-DNA Interactions*, Advances in Biochemical Engineering Biotechnology, Springer, 2007.

[3] H. Pearson, “What is a gene?” *Nature*, vol. 441, no. 7092, pp. 398–401, 2006.

[4] UCSC Genome Bioinformatics, http://hgdownload.cse.ucsc.edu/downloads.html.

[5] G. E. Sims, S. R. Jun, G. A. Wu, and S. H. Kim, “Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 8, pp. 2677–2682, 2009.

[6] W. J. Murphy, T. H. Pringle, T. A. Crider, M. S. Springer, and W. Miller, “Using genomic data to unravel the root of the placental mammal phylogeny,” *Genome Research*, vol. 17, no. 4, pp. 413–421, 2007.

[7] H. Zhao and G. Bourque, “Recovering genome rearrangements in the mammalian phylogeny,” *Genome Research*, vol. 19, no. 5, pp. 934–942, 2009.

[8] A. B. Prasad, M. W. Allard, and E. D. Green, “Confirming the phylogeny of mammals by use of large comparative sequence data sets,” *Molecular Biology and Evolution*, vol. 25, no. 9, pp. 1795–1808, 2008.

[9] I. Ebersberger, P. Galgoczy, S. Taudien, S. Taenzer, M. Platzer, and A. Von Haeseler, “Mapping human genetic ancestry,” *Molecular Biology and Evolution*, vol. 24, no. 10, pp. 2266–2276, 2007.

[10] C. W. Dunn, A. Hejnol, D. Q. Matus et al., “Broad phylogenomic sampling improves resolution of the animal tree of life,” *Nature*, vol. 452, no. 7188, pp. 745–749, 2008.

[11] J. A. T. Machado, A. C. Costa, and M. D. Quelhas, “Fractional dynamics in DNA,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, no. 8, pp. 2963–2969, 2011.

[12] A. M. Costa, J. T. Machado, and M. D. Quelhas, “Histogram-based DNA analysis for the visualization of chromosome, genome and species information,” *Bioinformatics*, vol. 27, no. 9, pp. 1207–1214, 2011.

[13] J. A. T. Machado, A. C. Costa, and M. D. Quelhas, “Entropy analysis of the DNA code dynamics in human chromosomes,” *Computers & Mathematics with Applications*, vol. 62, no. 3, pp. 1612–1617, 2011.

[14] J. A. T. Machado, A. C. Costa, and M. D. Quelhas, “Analysis and visualization of chromosome information,” *Gene*, vol. 491, no. 1, pp. 81–87, 2012.

[15] M. Kimura, *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge, Mass, USA, 1983.

[16] P. J. Deschavanne, A. Giron, J. Vilain, G. Fagot, and B. Fertit, “Genomic signature: characterization and classification of species assessed by chaos game representation of sequences,” *Molecular Biology and Evolution*, vol. 16, no. 10, pp. 1391–1399, 1999.

[17] M. Lynch, “The frailty of adaptive hypotheses for the origins of organismal complexity,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 1, pp. 8597–8604, 2007.

[18] G. Albrecht-Buehler, “Asymptotically increasing compliance of genomes with Chargaff’s second parity rules through inversions and inverted transpositions,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 47, pp. 17828–17833, 2006.

[19] D. Mitchell and R. Bridge, “A test of Chargaff’s second rule,” *Biochemical and Biophysical Research Communications*, vol. 340, no. 1, pp. 90–94, 2006.

[20] B. R. Powdel, S. S. Satapathy, A. Kumar et al., “A study in entire chromosomes of violations of the intra-strand parity of complementary nucleotides (Chargaff’s Second Parity Rule),” *DNA Research*, vol. 16, no. 6, pp. 325–343, 2009.

[21] C. T. Zhang, R. Zhang, and H. Y. Ou, “The Z curve database: a graphic representation of genome sequences,” *Bioinformatics*, vol. 19, no. 5, pp. 593–599, 2003.

[22] P. Bak, K. Chen, and C. Tang, “A forest-fire model and some thoughts on turbulence,” *Physics Letters A*, vol. 147, no. 5-6, pp. 297–300, 1990.

[23] N. E. Israeloff, M. Kagalenko, and K. Chan, “Can Zipf distinguish language from noise in noncoding DNA?” *Physical Review Letters*, vol. 76, pp. 1976–1979, 1995.

[24] R. N. Mantegna and H. E. Stanley, “Scaling behaviour in the dynamics of an economic index,” *Nature*, vol. 376, no. 6535, pp. 46–49, 1995.

[25] L. A. Adamic and B. A. Huberman, “Zipfs law and the Internet,” *Glottometrics*, vol. 3, pp. 143–150, 2002.

[26] H. Aoyama, Y. Fujiwara, and W. Souma, “Kinematics and dynamics of pareto-zipf’s law and gibrat’s law,” *Physica A*, vol. 344, no. 1-2, pp. 117–121, 2004.

[27] C. Andersson, A. Hellervik, and K. Lindgren, “A spatial network explanation for a hierarchy of urban power laws,” *Physica A*, vol. 345, no. 1-2, pp. 227–244, 2005.

[28] A. L. Barabási, “The origin of bursts and heavy tails in human dynamics,” *Nature*, vol. 435, no. 7039, pp. 207–211, 2005.

[29] W. Dahui, L. Menghui, and D. Zengru, “True reason for Zipf’s law in language,” *Physica A*, vol. 358, no. 2–4, pp. 545–550, 2005.

[30] J. M. Sarabia and F. Prieto, “The Pareto-positive stable distribution: a new descriptive model for city size data,” *Physica A*, vol. 388, no. 19, pp. 4179–4191, 2009.

[31] T. Fenner, M. Levene, and G. Loizou, “Predicting the long tail of book sales: unearthing the power-law exponent,” *Physica A*, vol. 389, no. 12, pp. 2416–2421, 2010.

[32] J. A. T. Machado, A. C. Costa, and M. D. Quelhas, “Shannon, Rényie and Tsallis entropy analysis of DNA using phase plane,” *Nonlinear Analysis: Real World Applications*, vol. 12, no. 6, pp. 3135–3144, 2011.

[33] J. A. T. Machado and S. Entropy, “Analysis of the Genome Code,” *Mathematical Problems in Engineering*, vol. 2012, Article ID 132625, 12 pages, 2012.

[34] J. T. Machado, “Accessing complexity from genome information,” *Communications in Nonlinear Science and Numerical Simulations*, vol. 17, no. 6, pp. 2237–2243, 2012.

[35] R. Hilfer, *Applications of Fractional Calculus in Physics*, World Scientific, Singapore, 2000.

[36] D. Baleanu and S. I. Vacaru, “Fractional curve flows and solitonic hierarchies in gravity and geometric mechanics,” *Journal of Mathematical Physics*, vol. 52, no. 5, Article ID 053514, 15 pages, 2011.

[37] D. Baleanu, K. Diethelm, E. Scalas, and J. J. Trujillo, *Fractional Calculus Models and Numerical Methods*, vol. 3 of *Complexity, Nonlinearity and Chaos*, World Scientific Publishing, 2012.

[38] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[39] E. T. Jaynes, “Information Theory and Statistical Mechanics,” vol. 106, pp. 620–630, 1957.

[40] A. I. Khinchin, *Mathematical foundations of information theory*, Dover Publications, New York, NY, USA, 1957.

[41] A. Plastino and A. R. Plastino, “Tsallis Entropy and Jaynes’ information theory formalism,” *Brazilian Journal of Physics*, vol. 29, no. 1, pp. 50–60, 1999.

[42] H. J. Haubold, A. M. Mathai, and R. K. Saxena, “Boltzmann-Gibbs entropy versus Tsallis entropy: recent contributions to resolving the argument of Einstein concerning “neither Herr Boltzmann nor Herr Planck has given a definition of W”? Essay review,” *Astrophysics and Space Science*, vol. 290, no. 3-4, pp. 241–245, 2004.

[43] A. M. Mathai and H. J. Haubold, “Pathway model, superstatistics, Tsallis statistics, and a generalized measure of entropy,” *Physica A*, vol. 375, no. 1, pp. 110–122, 2007.

[44] T. Carter, *An Introduction to Information Theory and Entropy*, Complex Systems Summer School, Santa Fe, Mexico, 2007.

[45] P. N. Rathie and S. Da Silva, “Shannon, Lévy, and Tsallis: a note,” *Applied Mathematical Sciences*, vol. 2, no. 25–28, pp. 1359–1363, 2008.

[46] C. Beck, “Generalised information and entropy measures in physics,” *Contemporary Physics*, vol. 50, no. 4, pp. 495–510, 2009.

[47] I. J. Taneja, "On measures of information and inaccuracy," *Journal of Statistical Physics*, vol. 14, no. 3, pp. 263–270, 1976.

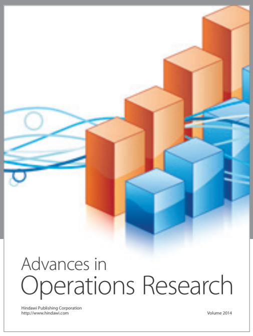
[48] B. D. Sharma and I. J. Taneja, "Three generalized-additive measures of entropy," *Elektronische Informationsverarbeitung und Kybernetik*, vol. 13, no. 7-8, pp. 419–433, 1977.

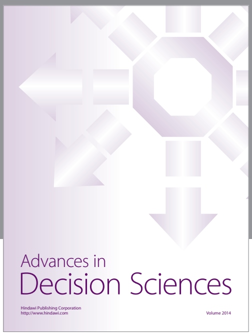
[49] A. Wehrl, "General properties of entropy," *Reviews of Modern Physics*, vol. 50, no. 2, pp. 221–260, 1978.

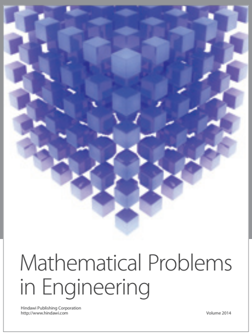
[50] H. D. Chen, C. H. Chang, L. C. Hsieh, and H. C. Lee, "Divergence and Shannon information in genomes," *Physical Review Letters*, vol. 94, no. 17, Article ID 178103, 2005.

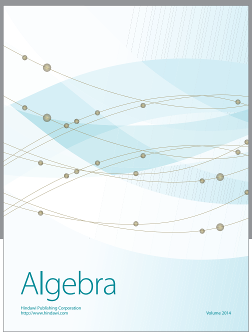
[51] R. M. Gray, *Entropy and Information Theory*, Springer, New York, NY, USA, 1990.

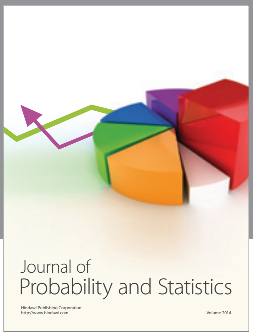
[52] M. R. Ubriaco, "Entropies based on fractional calculus," *Physics Letters A*, vol. 373, no. 30, pp. 2516–2519, 2009.


Advances in
Operations Research

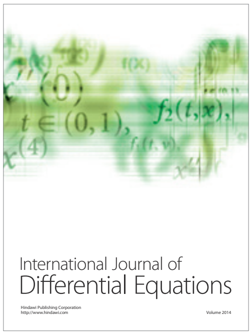

Advances in
Decision Sciences

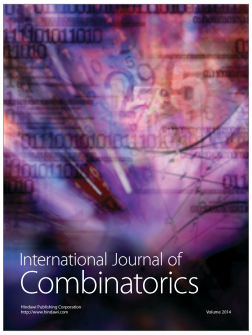

Mathematical Problems
in Engineering

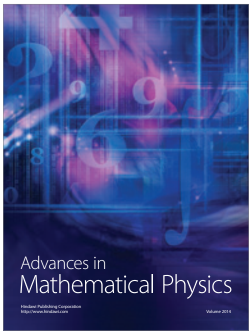

Algebra

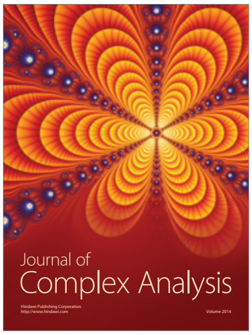

Journal of
Probability and Statistics

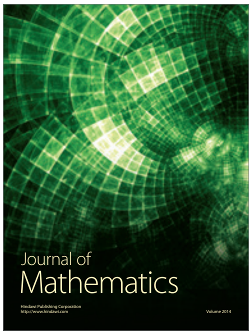
The Scientific
World Journal

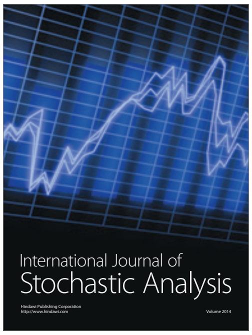

International Journal of
Differential Equations

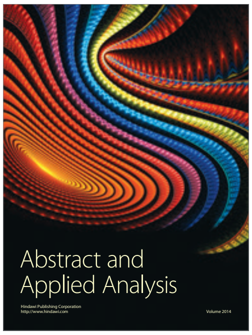

International Journal of
Combinatorics

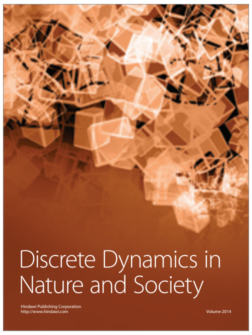

Advances in
Mathematical Physics

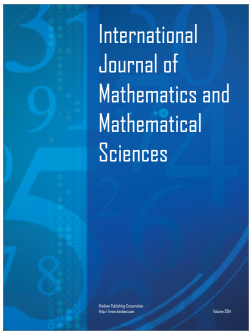

Journal of
Complex Analysis


Journal of
Mathematics


International Journal of
Stochastic Analysis

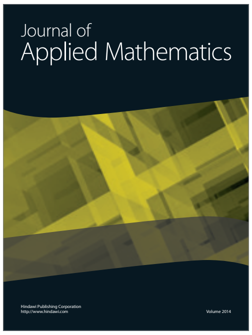

Abstract and
Applied Analysis


Discrete Dynamics in
Nature and Society


International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces


Journal of
Applied Mathematics

Journal of
Optimization