

Quantumlike Chaos in the Frequency Distributions of Bases A, C, G, T in Human Chromosome1 DNA

A. M. SELVAM

Deputy Director (Retired)

Indian Institute of Tropical Meteorology, Pune 411 008, INDIA

Abstract: - Recent studies of DNA sequence of letters A, C, G and T exhibit the inverse power law form frequency spectrum. Inverse power-law form of the power spectra of fractal space-time fluctuations is generic to the dynamical systems in nature and is identified as self-organized criticality. In this study it is shown that the power spectra of the frequency distributions of bases A, C, G, T in the Human chromosome 1 DNA exhibit self-organized criticality. DNA is a quasicrystal possessing maximum packing efficiency in a hierarchy of spirals or loops. Self-organized criticality implies that non-coding *introns* may not be redundant, but serve to organize the effective functioning of the coding *exons* in the DNA molecule as a complete unit.

Keywords: long-range correlations in DNA, self-organized criticality, quantumlike chaos, fractals, inverse power law, nonlocal connections

1. Introduction

DNA topology is of fundamental importance for a wide range of biological processes[1]. Since the topological state of genomic DNA is of importance for its replication, recombination and transcription, there is an immediate interest to obtain information about the supercoiled state from sequence periodicities[2,3]. Identification of dominant periodicities in DNA sequence will help understand the important role of coherent structures in genome sequence organization[4,5]. Li[6] has discussed meaningful applications of spectral analyses in DNA sequence studies. Recent studies indicate that the DNA sequence of letters A, C, G and T exhibit the inverse power law form $1/f^\alpha$ frequency spectrum where f is the frequency and α the exponent. It is possible, therefore, that the sequences have long-range order[7-14]. Inverse power-law form for power spectra of fractal space-time fluctuations is generic to dynamical systems in nature and is identified as self-organized criticality[15]. A general systems theory developed by the author[16-18] predicts the following: (a) long-range correlations intrinsic to self-organized criticality are a signature of quantumlike chaos, (b) the power spectra follow the universal inverse power law form of the statistical normal distribution, (c) fractal fluctuations self-organize to form the quasicrystalline structure of the space filling quasiperiodic Penrose tiling pattern. Here it is shown that power spectra of frequency distributions of bases A, C, G, T in Human chromosome 1 DNA exhibit self-organized criticality. DNA is a quasicrystal possessing maximum packing efficiency[19] in a hierarchy of spirals or loops.

2. Model Concepts

Power spectra of fractal space-time fluctuations of dynamical systems such as fluid flows, stock market price fluctuations, heart beat patterns, etc., exhibit inverse power-law form identified as self-organized criticality[15] and represent a selfsimilar eddy continuum. A general systems theory[16-18] developed by the author shows that such an eddy continuum can be visualised as a hierarchy of successively larger scale eddies enclosing smaller scale eddies. Since the large eddy is the integrated mean of the enclosed smaller eddies, the eddy energy (variance) spectrum follows the statistical normal distribution according to the Central Limit Theorem[20]. Hence the additive amplitudes of eddies, when squared, represent the probabilities, which is also an observed feature of the subatomic dynamics of quantum systems such as the electron or photon[21-23]. The long-range correlations intrinsic to self-organized criticality in dynamical systems are signatures of quantumlike chaos associated with the following characteristics: (a) The fractal fluctuations result from an overall logarithmic spiral trajectory with the quasiperiodic Penrose tiling pattern[16-18] for the internal

structure. (b) Conventional continuous periodogram power spectral analyses of such spiral trajectories will reveal a continuum of wavelengths with progressive increase in phase. (c) The broadband power spectrum will have embedded dominant wavebands, the bandwidth increasing with wavelength, and the wavelengths being functions of the golden mean. The first 13 values of the model predicted [16-18] dominant peak wavelengths are 2.2, 3.6, 5.8, 9.5, 15.3, 24.8, 40.1, 64.9, 105.0, 167.0, 275, 445.0 and 720 in units of the block length 10bp (base pairs). Wavelengths (or periodicities) close to the model predicted values have been reported in weather and climate variability [17], prime number distribution [24], Riemann zeta zeros (non-trivial) distribution [25], stock market economics [26]. (d) The conventional power spectrum plotted as the variance versus the frequency in log-log scale will now represent the eddy probability density on logarithmic scale versus the standard deviation of the eddy fluctuations on linear scale since the logarithm of the eddy wavelength represents the standard deviation, i.e. the r.m.s (root mean square) value of the eddy fluctuations. The r.m.s. value of the eddy fluctuations can be represented in terms of statistical normal distribution as follows. A normalized standard deviation $t=0$ corresponds to cumulative percentage probability density equal to 50 for the mean value of the distribution. For the overall logarithmic spiral circulation the logarithm of the wavelength represents the r.m.s. value of eddy fluctuations and the normalized standard deviation t is defined for the eddy energy as

$$t = \frac{\log_e L}{\log_e T_{50}} - 1 \quad (1)$$

The parameter L in Eq. 1 is the wavelength and T_{50} is the wavelength up to which the cumulative percentage contribution to total variance is equal to 50 and $t = 0$. The variable $\log_e T_{50}$ also represents the mean value for the r.m.s. eddy fluctuations and is consistent with the concept of the mean level represented by r.m.s. eddy fluctuations. Spectra of time series of fluctuations of dynamical systems, for example, meteorological parameters, when plotted as cumulative percentage contribution to total variance versus t follow the model predicted universal spectrum [17].

3. Data and analyses

The Human chromosome 1 DNA base sequence was obtained from the entrez Databases (Benson, Karsch-Mizrachi, Lipman, Ostell, Rapp and Wheeler, GenBank, Nucleic Acids Res. 30, 17-20 (2002)), Homo sapiens Genome (build 30) at <http://www.ncbi.nlm.nih.gov/entrez>. The first 10 contiguous data sets (Table 1) consisting of a total number of 9931745 bases were scanned to give a total number of 280 unbroken data sets of length 35000 bases each for the study. The number of times that each of the four bases A, C, G, T occur in successive blocks of 10 bases were determined giving 4 groups of 3500 frequency sequence values for each data set.

Table 1: Details of data sets

s.no	Accession number	Base pairs	
		from	to
1	NT_004350.12	1	2112206
2	NT_004321.12	1	1131467
3	NT_004547.12	1	92989
4	NT_034374.1	1	175802
5	NT_022058.11	1	156602
6	NT_021917.12	1	303717
7	NT_028054.9	1	3621417
8	NT_032967.2	1	80421
9	NT_021937.12	1	2077871
10	NT_034375.1	1	179253

The power spectra of frequency distribution of bases were computed accurately by an elementary, but very powerful method of analysis developed by Jenkinson[27] which provides a quasi-continuous form of the classical periodogram allowing systematic allocation of the total variance and degrees of freedom of the data series to logarithmically spaced elements of the frequency range (0.5, 0). The cumulative percentage contribution to total variance was computed starting from the high frequency side of the spectrum. The power spectra were plotted as cumulative percentage contribution to total variance versus the normalized standard deviation t . The average variance spectra for the 280 data sets and the statistical normal distribution are shown in Fig. 1 for the four bases. The 'goodness of fit' (statistical chi-square test) between the variance spectra and statistical normal distribution is significant at less than or equal to 5% level for 98.6, 99.3, 98.9, 97.9 percent of the 280 data sets respectively for the four bases A, C, G and T. The average and standard deviation of the wavelength T_{50} up to which the cumulative percentage contribution to total variance is equal to 50 are also shown in Fig. 1. The power spectra exhibit dominant wavebands where the normalized variance is equal to or greater than 1. The dominant peak wavelengths were grouped into 13 class intervals 2 - 3, 3 - 4, 4 - 6, 6 - 12, 12 - 20, 20 - 30, 30 - 50, 50 - 80, 80 - 120, 120 - 200, 200 - 300, 300 - 600, 600 - 1000 (in units of 10bp block lengths) to include the model predicted dominant peak length scales mentioned earlier. Average class interval-wise percentage frequencies of occurrence of dominant wavelengths are shown in Fig. 2 along with the percentage contribution to total variance in each class interval corresponding to the normalised standard deviation t computed from the average T_{50} (Fig. 1) for each of the four bases.

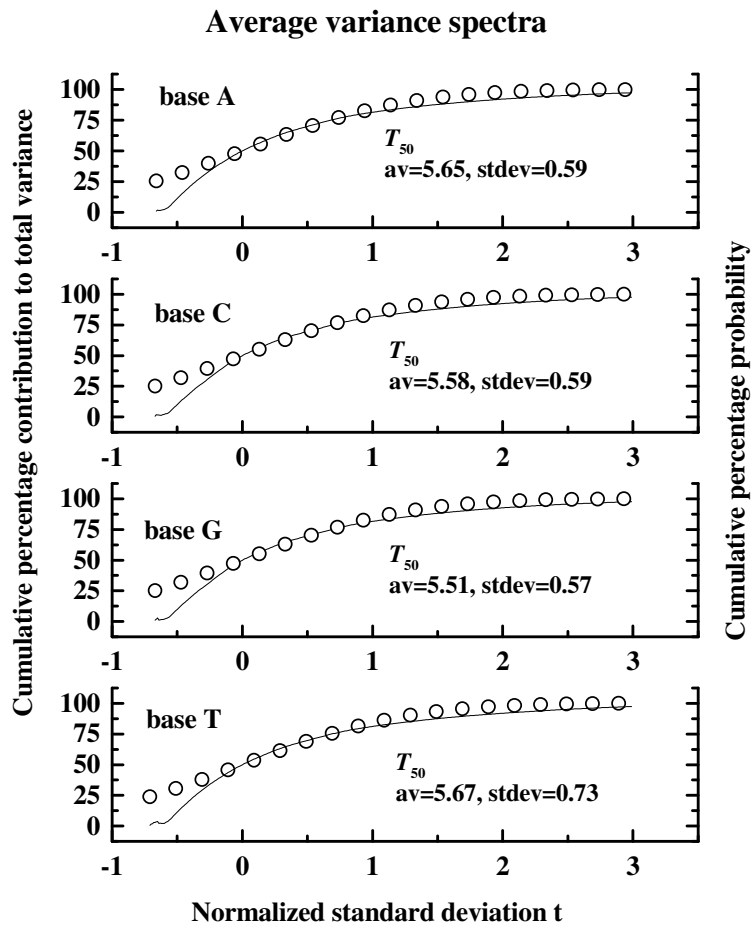


Figure 1: Average variance spectra for the four bases in Human chromosome 1 DNA. Continuous lines are for the variance spectra and open circles give the statistical normal distribution. The mean and standard deviation of the wavelengths T_{50} up to which the cumulative percentage contribution to total variance is equal to 50 are also given in the figure.

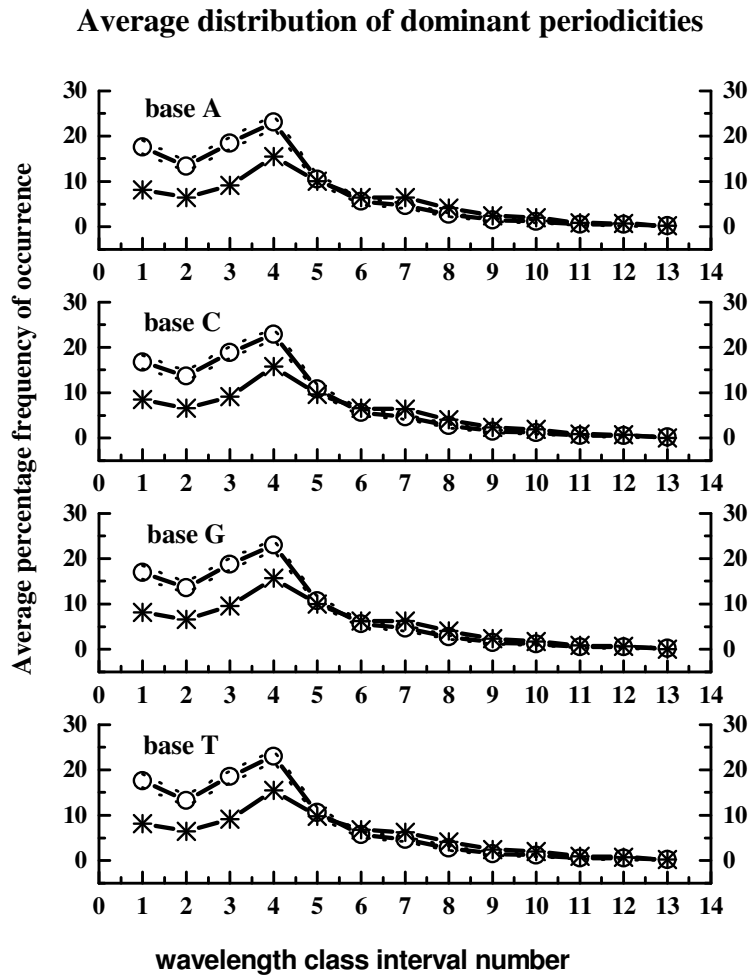


Figure 2: Average wavelength class interval-wise percentage distribution of dominant (normalized variance greater than 1) wavelengths. Line + open circle give the average and dotted lines denote one standard deviation on either side of the mean. The computed percentage contribution to the total variance for each class interval is given by line + star.

4. Results and conclusions

The variance spectra for almost all the 280 data sets exhibit the universal inverse power-law form $1/f^\alpha$ of the statistical normal distribution (Fig. 1) where f is the frequency and the spectral slope α decreases with increase in wavelength and approaches 1 for long wavelengths. The above result is also seen in Fig. 2 where the wavelength class interval-wise percentage frequency distribution of dominant wavelengths follow closely the corresponding computed variation of percentage contribution to the total variance as given by the statistical normal distribution. Inverse power-law form for power spectra implies long-range spatial correlations in the frequency distributions of the bases in DNA. Microscopic-scale quantum systems such as the electron or photon exhibit non-local connections or long-range correlations and are visualized to result from the superimposition of a continuum of eddies. Therefore, by analogy, the observed fractal fluctuations of the frequency distributions of the bases exhibit quantumlike chaos in the Human chromosome 1 DNA. The eddy continuum acts as a robust unified whole fuzzy logic network with global response to local perturbations. Therefore, artificial modification of base sequence structure at any location may have significant noticeable effects on the function of the DNA molecule as a whole. Further, the presence of introns, which do not have

meaningful code, may not be redundant, but may serve to organize the effective functioning of the coding exons in the DNA molecule as a complete unit[8].

The results imply that the DNA base sequence self-organizes spontaneously to generate the robust geometry of logarithmic spiral with the quasiperiodic Penrose tiling pattern for the internal structure. The space filling geometric figure of the Penrose tiling pattern has intrinsic local five-fold symmetry[28] and ten fold symmetry. One of the three basic components of DNA, the deoxyribose is a five-carbon sugar and may represent the local five-fold symmetry of the quasicrystalline structure of the quasiperiodic Penrose tiling pattern of the DNA molecule as a whole. The DNA molecule shows ten fold symmetry in the arrangement of 10 bases per turn of the double helix. The study of plant phyllotaxis in botany shows that quasicrystalline structure provides maximum packing efficiency for seeds, florets, leaves, etc[29,19,30]. Quasicrystalline structure of the quasiperiodic *Penrose* tiling pattern may be the geometrical structure underlying the packing of 10^3 to 10^5 micrometer of DNA in a eukaryotic (higher organism) chromosome into a metaphase structure a few microns long. The spatial geometry of the DNA is therefore organized into a hierarchy of helical structures. Such a concept may explain the observed loops of DNA in metaphase chromosome[31]. For example, the average class-interval wise percentage distribution of dominant periodicities show a peak in the wavelength interval 6-12 in units of 10bp, i.e. 60 to 120bp for all the four bases (Fig. 2). This predominant wavelength interval 60 to 120bp may correspond to the coil length of each of the two DNA coils on the basic nucleosome unit of the chromatin fibre. Also, the value of T_{50} ranges from 5 to 6 in units of 10bp, i.e. from 50 to 60bp (Fig. 1) indicating again the predominance of the fundamental coil length in the double coil of DNA in nucleosomes.

Acknowledgement

The author is grateful to Dr. A. S. R. Murty for encouragement.

References

- [1] Bates, A. D. & Maxwell, A. *DNA Topology*. Oxford University Press, Oxford, pp.111 (1993).
- [2] Herzel, H., Weiss, O., & Trifonov, E. N. Sequence periodicity in complete genomes of Archaea suggests positive supercoiling. *Journal of Biomolecular Structure & Dynamics* **16(2)**, 341-345 (1998). http://linkage.rockefeller.edu/wli/dna_corr/1998.html
- [3] Herzel, H., Weiss, O., & Trifonov, E. N. 10-11 bp periodicities in complete genomes reflect protein structure and DNA folding. *Bioinformatics* **15(3)**, 187-193 (1999). http://linkage.rockefeller.edu/wli/dna_corr/1999.html
- [4] Chechetkin, V. R. and Yu. Turygin, A. Search of hidden periodicities in DNA sequences. *Journal of Theoretical Biology* **175**, 477-494 (1995). http://linkage.rockefeller.edu/wli/dna_corr/1995.html
- [5] Widom, J. Short-range order in two eukaryotic genomes: relation to chromosome structure. *Journal of Molecular Biology* **259**, 579-588 (1996). http://linkage.rockefeller.edu/wli/dna_corr/widom96.pdf
- [6] Li, W. Are spectral analyses useful for DNA sequence analysis? *Proc. DNA in Chromatin, At the Frontiers of Biology, Biophysics, and Genomics*, March 23-29 (2002). Arcachon, France. <http://linkage.rockefeller.edu/wli/pub/arcachon02.pdf>
- [7] Li, W. Generating nontrivial long-range correlations and 1/f spectra by replication and mutation. *International Journal of Bifurcation and Chaos* **2(1)**, 137-154 (1992). http://linkage.rockefeller.edu/wli/dna_corr/1-ijbc92-1.html
- [8] Selvam, A. M. Quantumlike chaos in the frequency distributions of the bases A, C, G, T in *Drosophila* DNA. *APEIRON* **9(4)**, 103-148 (2002). <http://redshift.vif.com/JournalFiles/V09NO4PDF/V09N4sel.pdf>

- [9] Li, W., Marr, T. G., Kaneko, K. Understanding long-range correlations in DNA sequences. *Physica D* **75(1-3)**, 392-416 (1994); erratum: **82**, 217 (1995). <http://arxiv.org/chao-dyn/9403002>
- [10] Audit, B. *et al.* Long-range correlations in genomic DNA: a signature of the nucleosomal structure. *Phys. Rev. Lett.* **86(11)**, 2471-2474 (2001). http://linkage.rockefeller.edu/wli/dna_corr/audit01.pdf
- [11] Stanley H. E., Amaral, L. A. N., Gopikrishnan, P., and Plerou, V. Scale invariance and universality of economic fluctuations. *Physica A* **283**, 31-41 (2000).
- [12] Stanley, H. E. Exotic statistical physics: applications to biology, medicine, and economics. *Physica A* **285**, 1-17 (2000). <http://www.elsevier.com/gej-ng/10/36/21/81/30/28/article.pdf>
- [13] Audit, B., Vaillant, C., Arneodo, A., d'Aubenton-Carafa, Y., Thermes, C. Long-range correlations between DNA bending sites: relation to the structure and dynamics of nucleosomes. *Journal of Molecular Biology* **316(4)**, 903-918 (2002).
- [14] Som, A. Chattopadhyay, Chakrabarti, J. and Bandyopadhyay, D. Codon distributions in DNA. *Phys. Rev. E* **63**, 1-8 (2001). http://linkage.rockefeller.edu/wli/dna_corr/som01.pdf
- [15] Bak, P., Tang, C., Wiesenfeld, K. Self-organized criticality: an explanation of $1/f$ noise. *Phys. Rev. Lett.* **59**, 381-384 (1987).
- [16] Mary Selvam, A. Deterministic chaos, fractals and quantumlike mechanics in atmospheric flows. *Can. J. Phys.* **68**, 831-841 (1990). <http://xxx.lanl.gov/html/physics/0010046>
- [17] Selvam, A. M., and Fadnavis, S. Signatures of a universal spectrum for atmospheric interannual variability in some disparate climatic regimes. *Meteorol. & Atmos. Phys.* **66**, 87-112 (1998). <http://xxx.lanl.gov/abs/chao-dyn/9805028>
- [18] Selvam, A. M. and Suvarna Fadnavis. Superstrings, cantorion-fractal space-time and quantum-like chaos in atmospheric flows. *Chaos, Solitons and Fractals* **10(8)**, 1321-1334 (1999). <http://xxx.lanl.gov/abs/chao-dyn/9806002>
- [19] Stewart, I. Daisy, daisy, give your answer do. *Sci. Amer.* **272**, 76-79 (1995).
- [20] Ruhla, C. *The Physics of Chance*. Oxford University Press, Oxford, pp.217 (1992).
- [21] Maddox, J. Licence to slang Copenhagen? *Nature* **332**, 581 (1988).
- [22] Maddox, J. Can quantum theory be understood? *Nature* **361**, 493 (1993).
- [23] Rae, A. *Quantum-physics: illusion or reality?* Cambridge University Press, New York, pp.129 (1988).
- [24] Selvam, A. M. Quantumlike chaos in prime number distribution and in turbulent fluid flows. *APEIRON* **8(3)**, 29-64 (2001). <http://redshift.vif.com/JournalFiles/V08NO3PDF/V08N3SEL.PDF>
<http://xxx.lanl.gov/html/physics/0005067>
- [25] Selvam, A. M. Signatures of quantumlike chaos in spacing intervals of non-trivial Riemann zeta zeros and in turbulent fluid flows. *APEIRON* **8(4)**, 10-40 (2001). <http://redshift.vif.com/JournalFiles/V08NO4PDF/V08N4SEL.PDF>
<http://xxx.lanl.gov/html/physics/0102028>
- [26] Sornette, D., Johansen, A., & Bouchaud, J-P. Stock market crashes, precursors and replicas. <http://xxx.lanl.gov/pdf/cond-mat/9510036> (1995).
- [27] Jenkinson, A. F. *A Powerful Elementary Method of Spectral Analysis for use with Monthly, Seasonal or Annual Meteorological Time Series*. Meteorological Office, London, Branch Memorandum No. 57, pp. 1-23 (1997).

- [28] Devlin, K. *Mathematics: The Science of Patterns*. Scientific American Library, NY, p.101 (1997).
- [29] Jean R. V. *Phyllotaxis: A Systemic Study in Plant Morphogenesis*. Cambridge University Press, NY, USA (1994).
- [30] Mary Selvam, A. Quasicrystalline pattern formation in fluid substrates and phyllotaxis. In *Symmetry in Plants*, D. Barabe and R. V. Jean (Editors), World Scientific Series in Mathematical Biology and Medicine, Volume 4., Singapore, pp.795-809 (1998). <http://xxx.lanl.gov/abs/chaodyn/9806001>
- [31] Grosveld, F. and Fraser, P. Locus control of regions. In *Nuclear Organization, Chromatin Structure, and Gene Expression*. pp. 129-144. (eds.) Roel Van Driel and Arie P Otte, Oxford University Press (1997).