Universal Spectrum for DNA Base C+G Concentration Variability in Human Chromosome Y

A. M. SELVAM
Deputy Director (Retired)
Indian Institute of Tropical Meteorology, Pune 411 008, INDIA
http://www.geocities.com/amselvam
http://amselvam.tripod.com/index.html

Abstract: - The spatial distribution of DNA base sequence A, C, G and T exhibit selfsimilar fractal fluctuations and the corresponding power spectra follow inverse power law of the form $1/f^{\alpha}$ where f is the frequency and α the exponent. Inverse power law form for power spectra implies long-range spatial correlations of the eddy fluctuations. Multifractal structure to space-time fluctuations and the associated inverse power law form for power spectra is generic to spatially extended dynamical systems in nature and is a signature of self-organized criticality. The exact physical mechanism for the observed self-organized criticality is not yet identified. The author has developed a general systems theory where quantum mechanical laws emerge as self-consistent explanations for the observed long-range space-time correlations, i.e. the apparently chaotic fractal fluctuations are signatures of quantum-like chaos in dynamical systems. The model also provides unique quantification for the observed inverse power law form for power spectra in terms of the statistical normal distribution. In this paper it is shown that the frequency distribution of the bases C+G in all available contiguous sequences for Human chromosome Y DNA exhibit model predicted quantum-like chaos.

Key-Words: - fractals, chaos, self-organized criticality, quasicrystalline structure, quantumlike chaos

1 Introduction

Long-range space-time correlations, manifested as the selfsimilar fractal geometry to the spatial pattern, concomitant with inverse power law form for power spectra of space-time fluctuations are generic to spatially extended dynamical systems in nature and are identified as signatures of self-organized criticality. Self-organized criticality implies non-local connections in space and time, i.e., long-term memory of short-term spatial fluctuations in the extended dynamical system that acts as a unified whole communicating network. 'Nonlinear dynamics and chaos', a multidisciplinary area of intensive research in recent years (since 1980s) has helped identify universal characteristics of spatial patterns (forms) and temporal fluctuations (functions) of disparate dynamical systems in nature [1]. Examples of dynamical systems, i.e., systems which change with time include biological (living) neural networks of the human brain which responds as a unified whole to a multitude of input signals and the non-biological (non-living) atmospheric flow structure which exhibits teleconnections, i.e., long-range space-time correlations. Spatially extended dynamical systems in nature exhibit selfsimilar fractal geometry to the spatial pattern. The fractal structure of physiological systems has been identified [2, 3]. Fractal architecture to the spatial pattern enables integration of a multitude of signals of all space-time scales so that the dynamical system responds as a unified whole to local stimuli, i.e., short term fluctuations are carried as internal structure to long-term fluctuations. Power spectral analysis [4] is conventionally used to resolve the periodicities (frequencies) and their amplitudes in time series data of fluctuations. The power spectrum is plotted on log-log scale as the intensity represented by variance (amplitude squared) versus the period (frequency) of the component periodicities. Dynamical systems in nature exhibit inverse power law form $1/f^{\alpha}$ where f is the frequency (1/period) and α the exponent for the power spectra of space-time fluctuations indicating selfsimilar fluctuations on all space-time scales. The universal characteristics of spatially extended dynamical systems, namely, the fractal structure to the space-time fluctuation pattern and inverse power law form for power spectra of space-time fluctuations are identified as signatures of self-organized criticality [5]. The physical mechanisms responsible for self-organized criticality should be independent of the exact details (physical, chemical, physiological, biological, computational system etc.) of the dynamical system so as to be universally applicable to all dynamical systems (real and model).

Atmospheric flows exhibit self-organized criticality as manifested in the fractal geometry to the global cloud cover pattern concomitant with inverse power law form for power spectra of temporal fluctuations of meteorological parameters. Standard models for atmospheric flow dynamics cannot explain the observed self-organized criticality in atmospheric flows satisfactorily. The author has developed a general systems theory for atmospheric flows [6, 7] that predicts the observed self-organized criticality as intrinsic to quantumlike chaos governing flow dynamics. In the following Section 2, the model for self-organized criticality in atmospheric flows is first summarized and model concepts are shown to be applicable to all real world and model dynamical systems. The concept of self-organized criticality and quantumlike chaos in biological and physiological systems in particular are discussed.

2 General systems theory concepts

In summary [6, 7], the model is based on Townsend's concept [8] that large eddy structures form in turbulent flows as envelopes of enclosed turbulent eddies. Such a simple concept that space-time averaging of small-scale structures gives rise to large-scale space-time fluctuations leads to the following important model predictions. Since the large eddy is but the integrated mean of enclosed turbulent eddies, the eddy energy (kinetic) distribution follows statistical normal distribution according to the Central Limit Theorem [9]. Such a result, that the additive amplitudes of the eddies, when squared, represent probability distributions is found in the subatomic dynamics of quantum systems such as the electron or photon. Atmospheric flows, or in general turbulent fluid flows follow quantumlike chaos.

The root mean square (r.m.s.) circulation speeds W and w_* of large and turbulent eddies of respective radii R and r are related as

$$W^2 = \frac{2}{\pi}\frac{r}{R}w_*^2 \qquad (1)$$

Equation (1) is a statement of the law of conservation of energy for eddy growth in fluid flows and implies a two-way ordered energy flow between the larger and smaller scales. Microscopic scale perturbations are carried permanently as internal circulations of progressively larger eddies. The flow structure consists of an overall logarithmic spiral trajectory with Fibonacci winding number and quasiperiodic Penrose tiling pattern for internal structure characterized by a nested continuum of vortices, i.e., vortices within vortices. The quasiperiodic Penrose tiling pattern with five-fold symmetry has been identified as quasicrystalline structure in condensed matter physics. Self-organized quasicrystalline pattern formation exists at the molecular level also and may result in condensation of specific biochemical structures in biological media. Logarithmic spiral formation with Fibonacci winding number and five-fold symmetry possess maximum packing efficiency for component parts and are manifested strikingly in *Phyllotaxis* [10].

Dominant quasi-periodicities P_n corresponding to the internal circulations (vortices) are given as

$$P_n = T(2+\tau)\tau^n \qquad (2)$$

The dominant quasi-periodicities are equal to *2.2T*, *3.6T*, *5.8T*, *9.5T*, ……for values of n = *-1*, *0*, *1*, *2*, respectively (Equation 2). Space-time integration of turbulent fluctuations results in robust broadband dominant periodicities which are functions of the primary perturbation time period *T* alone and are independent of exact details (chemical, electrical, physical etc.) of turbulent fluctuations. Wavelengths (or periodicities) close to the model predicted values have been reported in weather and climate variability [6, 7] prime number distribution [11], Riemann zeta zeros (non-trivial) distribution [12], Drosophila DNA base sequence [13], stock market economics [14], human chromosome 1 DNA base sequence [15].

Macroscale coherent structures emerge by space-time integration of microscopic domain fluctuations in fluid flows. Such a concept of the autonomous growth of atmospheric eddy continuum with ordered energy flow between the scales is analogous to Prigogine's ([16] concept of the spontaneous emergence of order and organization out of apparent disorder and chaos through a process of self-organization. General systems theory is a logical-mathematical field, the subject matter of which is the formulation and deduction of those principles which are valid for 'systems' in general, whatever the nature of their component elements or the relations or 'forces' between them ([17-19].

The logarithmic spiral flow pattern enclosing the component vortices may be visualized as a continuous smooth rotation of the phase angle θ with increase in period. The phase angle θ for each stage of growth is equal to *r/R* and is proportional to the variance W^2 (Equation 1), the variance representing the intensity of fluctuations. The phase angle gives a measure of coherence or correlation in space-time fluctuations. The model predicted continuous smooth rotation of phase angle with increase in period length associated with logarithmic spiral flow structure is analogous to Berry's phase [20] in quantum systems. Conventional power spectral analysis will resolve such a logarithmic spiral flow trajectory as a continuum of eddies (broadband spectrum) with a progressive increase in phase angle. The power spectrum, plotted on log-log scale as variance versus frequency (period) will represent the probability density corresponding to normalized standard deviation *t* given by

$$t = \frac{\log L}{\log T_{50}} - 1 \qquad (3)$$

In the above Equation (3) *L* is the period (or wavelength) and T_{50} is the period up to which the cumulative percentage contribution to total variance is equal to *50*. The above expression for normalized standard deviation *t* follows from model prediction of logarithmic spiral flow structure and model concept of successive growth structures by space-time averaging. Fluctuations of all scales therefore self-organize to form the universal inverse power law form of the statistical normal distribution. Since the phase angle θ equal to *r/R* represents the variance W^2 (Equation 1), the phase spectrum plotted similar to variance spectrum will also follow the statistical normal distribution.

The apparent paradoxes found in the subatomic dynamics of quantum systems [21] are consistent in the context of atmospheric flows as explained in the following. A quantum system behaves as a wave on some occasions and as a particle at other times. Wave-particle duality is consistent in the context of atmospheric waves, which generate particle-like clouds in a row because of formation of clouds in updrafts and dissipation of clouds in adjacent downdrafts characterizing wave motion. The separated parts of a quantum system respond as a unified whole to local perturbations. Non-local connection is implicit to atmospheric flow structure quantified in Equation (1) as ordered two-way energy flow between larger and smaller scales and seen as long-range space-time correlations, namely self-organized criticality. Atmospheric flows self-organize to form a unified network with the quasiperiodic Penrose tiling pattern for internal structure, which provide long-range (non-local) space-time connections.

3 Applications of the general systems theory concepts to genomic DNA base sequence structure

DNA sequences, the blueprint of all essential genetic information, are polymers consisting of two complementary strands of four types of bases: adenine (A), cytosine (C), guanine (G) and thymine (T). Among the four bases, the presence of A on one strand is always paired with T on the opposite strand, forming a "base pair" with 2 hydrogen bonds. Similarly, G and C are complementary to one another, while forming a base pair with 3 hydrogen bonds. Consequently, one may characterize AT base-pairs as weak bases and GC base-pairs as strong bases. In addition, the frequency of A(G) on a single strand is approximately equal to the frequency of T(C) on the same strand, a phenomenon that has been termed "strand symmetry" or "Chargaff's second parity". Therefore, DNA sequences can be transformed into sequences of weak W (A or T) and strong S (G or C) bases. The SW mapping rule is particularly appropriate to analyze genome-wide correlations; this rule corresponds to the most fundamental partitioning of the four bases into their natural pairs in the double helix (G+C, A+T). The composition of base pairs, or GC level, is thus a strand-independent property of a DNA molecule and is related to important physico-chemical properties of the chain [22]. The C+G content (isochore) studies have been done earlier [23-26]. The full story of how DNA really functions is not merely what is written on the sequence of base pairs. The DNA functions involve information transmission over many length scales ranging from a few to several hundred nanometers [27].

One of the major goals in DNA sequence analysis is to gain an understanding of the overall organization of the genome, in particular, to analyze the properties of the DNA string itself. Long-range correlations in DNA base sequence structure, which give rise to $1/f$ spectra have been identified [28, 29]. Such long-range correlations in space-time fluctuations is very common in nature and Li [30] has given an extensive and informative bibliography of the observed $1/f$ noise or $1/f$ spectra, where f is the frequency, in biological, physical, chemical and other dynamical systems. The long-range correlations in nucleotide sequence could in principle be explained by the coexistence of many different length scales. The advantage of spectral analysis is to reveal patterns hidden in a direct correlation function. The quality of the $1/f$ spectra differs greatly among sequences. Different DNA sequences do not exhibit the same power spectrum.

The concentration of genes is correlated with the C+G density. The spatial distribution of C+G density can be used to give an indication of the location of genes. The final goal is to eventually learn the 'genome organization principles' [31]. The coding sequences of most vertebrate genes are split into segments (exons), which are separated by noncoding intervening sequences (introns). A very small minority of human genes lack noncoding introns and are very small genes [32]. Li [33] reports that spectral analysis shows that there are GC content fluctuations at different length scales in isochore (relatively homogeneous) sequences. Fluctuations of all size scales coexist in a hierarchy of domains within domains [34]. Li and Holste [35] have recently identified universal $1/f$ spectra and diverse correlation structures in Guanine (G) and Cytosine (C) content of all human chromosomes.

4 Data and Analysis

4.1 Data

The Human chromosome Y DNA base sequence was obtained from the entrez Databases, Homo sapiens Genome (build 34 Version 2) at http://www.ncbi.nlm.nih.gov/entrez. The ten contiguous data sets, each containing a minimum of 70 000 base pairs, chosen for the study are given in Table 1.

4.2 Power spectral analyses: variance and phase spectra

The number of times base C and also base G, i.e., (C+G), occur in successive blocks of 10 bases were determined in successive length sections of 70000 base pairs giving a C+G frequency distribution series of 7000 values for each data set. The power spectra of frequency distribution of C+G bases (per 10bp) in the data sets were computed accurately by an elementary, but very powerful method of analysis developed by [36] 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 equal to ($\log L/\log T_{50}$) –1 where L is the period in years and T_{50} is the period up to which the cumulative percentage contribution to total variance is equal to *50* (Equation 3)). The corresponding phase spectra were computed as the cumulative percentage contribution to total rotation (Section 2). The statistical chi-square test [37] was applied to determine the 'goodness of fit' of variance and phase spectra with statistical normal distribution. Details of data sets and results of power spectral analyses are given in Table 1. The average variance and phase spectra for the data sets in each of the ten contiguous data series are given in Figure 1.

4.3 Power spectral analyses: dominant periodicities

The general systems theory predicts the broadband power spectrum of fractal fluctuations will have embedded dominant wavebands, the bandwidth increasing with wavelength, and the wavelengths being functions of the golden mean (Equation 2). The first 13 values of the model predicted [6, 7] 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) in the present study. 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 above. The class intervals increase in size progressively to accommodate model predicted increase in bandwidth associated with increasing wavelength. Average class interval-wise percentage frequencies of occurrence of dominant wavelengths (normalized variance greater than 1) are shown in Figure 2 along with the percentage contribution to total variance in each class interval corresponding to the normalised standard deviation t (Equation 3) computed from the average T_{50} (Table 1) for the ten data sets. In this context it may be mentioned that statistical normal probability density distribution represents the eddy variance (Equation 3). The observed frequency distribution of dominant eddies follow closely the computed percentage contribution to total variance.

5. Discussions

In summary, a majority of the data sets (Table 1 and Figure 1) exhibit the model predicted quantumlike chaos for fractal fluctuations since the variance and phase spectra follow each other closely and also follow the universal inverse power law form of the statistical normal distribution signifying long-range correlations or coherence in the overall frequency distribution pattern of the bases C+G in Human chromosome Y DNA. Such non-local connections or 'memory' in the spatial pattern is a natural consequence of the model predicted Fibonacci spiral enclosing the space filling quasicrystalline structure of the quasiperiodic Penrose tiling pattern for fractal fluctuations of dynamical systems. Further, the broadband power spectra exhibit dominant wavelengths closely corresponding to the model predicted (Equation 2, Figure 2) nested continuum of eddies. The apparently chaotic fluctuations of the frequency distribution of the bases C+G per 10bp in the Human chromosome Y DNA self-organize to form an ordered hierarchy of spirals or loops.

Quasicrystalline structure of the quasiperiodic Penrose tiling pattern has maximum packing efficiency as displayed in plant phyllotaxis [38] and 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 (before cell division) a few microns long as explained in the following. A length of DNA equal to $2\pi L$ when coiled in a loop of radius L has a packing efficiency (lengthwise) equal to $2\pi L/2L=\pi$ since the linear length $2\pi L$ is now accommodated in a length equal to the diameter $2L$ of the loop. Since each stage of looping gives a packing efficiency equal to π, ten stages of such successive looping will result in a packing efficiency equal to π^{10} approximately equal to 10^5.

The present study deals with Y chromosome bases C+G concentration per 10bp in all available contiguous sequences. The window length 10bp was chosen since the primary loop in the DNA molecule is equal to about 10bp. The power spectral analysis gives the dominant wavelengths in terms of this basic unit, namely the window length of 10bp. Increasing the window length (more than 10bp) will result in decrease in resolution of shorter wavelengths. The aim of this preliminary study is to determine the spatial organization of the DNA bases C+G by applying concepts of a general systems theory first developed for atmospheric flows. The author intends to do similar analyses and compare all human chromosomes and between species.

6. Conclusions

A recently developed cell dynamical system model for turbulent fluid flows predicts self-organized criticality as intrinsic to quantumlike mechanics governing flow dynamics. The model concepts are independent of exact details (physical, chemical, biological etc.) of the dynamical system and are universally applicable. The important conclusions of this study are as follows: (1) the frequency distribution of bases C+G per 10bp in human chromosome Y DNA exhibit selfsimilar fractal fluctuations which follow the universal inverse power law form of the statistical normal distribution (Figure 1), a signature of quantumlike chaos. (2) Quantumlike chaos indicates long-range spatial correlations or 'memory' inherent to the self-organized fuzzy logic network of the quasiperiodic Penrose tiling pattern (Equation 2). (3) Such non-local connections indicate that coding exons together with non-coding introns contribute to the effective functioning of the DNA molecule as a unified whole. Recent studies indicate that mutations in introns introduce adverse genetic defects [39]. (4) The space filling quasiperiodic Penrose tiling pattern provides maximum packing efficiency for the DNA molecule inside the chromosome.

Acknowledgement

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

References:

[1] Gleick, J., 1987. *Chaos: Making a New Science*, Viking, New York.

[2] Goldberger, A. L., Amaral, L. A. N., Hausdorff, J. M., Ivanov, P. Ch., Peng, C.- K., Stanley, H. E., 2002. Fractal dynamics in physiology: alterations with disease and aging. *PNAS* 99, Suppl.1, 2466-2472. http://pnas.org/cgi/doi/10.1073/pnas.012579499 .

[3] West, B. J., 2004. Comments on the renormalization group, scaling and measures of complexity. *Chaos, Solitons and Fractals* 20, 33-44.

[4] Macdonald, G. F., 1989. Spectral analysis of time series generated by nonlinear processes. *Reviews of Geophysics* 27, 449-469.

[5] Bak, P., Tang, C. and Wiesenfeld, K., 1988. Self-organized criticality. *Phys. Rev. A 38*, 364-374.

[6] Selvam, A. M., 1990. Deterministic chaos, fractals and quantumlike mechanics in atmospheric flows. *Canadian J. Physics* 68, 831-841. http://xxx.lanl.gov/html/physics/0010046

[7] Selvam, A. M., and Fadnavis, S., 1998. Signatures of a universal spectrum for atmospheric interannual variability in some disparate climatic regimes. *Meteorol. & Atmos. Phys.* 66, 87-112. http://xxx.lanl.gov/abs/chao-dyn/9805028

[8] Townsend, A. A., 1956. *The Structure of Turbulent Shear Flow*, second edition, U.K., Cambridge University Press.

[9] Ruhla, C., 1992. *The Physics of Chance*, Oxford University Press, Oxford, U. K., pp.217.

[10] Jean, R. V., 1994. Phyllotaxis: *A Systemic Study in Plant Morphogenesis*, Cambridge University Press, NY, USA.

[11] Selvam, A. M., 2001a. Quantumlike chaos in prime number distribution and in turbulent fluid flows. *APEIRON* 8(3), 29-64. http://redshift.vif.com/JournalFiles/V08NO3PDF/V08N3SEL.PDF http://xxx.lanl.gov/html/physics/0005067

[12] Selvam, A. M., 2001b. Signatures of quantumlike chaos in spacing intervals of non-trivial Riemann Zeta zeros and in turbulent fluid flows, *APEIRON* 8(4), 10-40. http://xxx.lanl.gov/html/physics/0102028 http://redshift.vif.com/JournalFiles/V08NO4PDF/V08N4SEL.PDF

[13] Selvam, A. M., 2002. Quantumlike chaos in the frequency distributions of the bases A, C, G, T in Drosophila DNA. *APEIRON* 9(4), 103-148. http://redshift.vif.com/JournalFiles/V09NO4PDF/V09N4sel.pdf

[14] Selvam, A. M., 2003. Signatures of quantum-like chaos in Dow Jones Index and turbulent fluid flows. http://arxiv.org/html/physics/0201006 *APEIRON* 10, 1-28. http://redshift.vif.com/JournalFiles/V10NO4PDF/V10N4SEL.PDF

[15] Selvam, A. M., 2004. Quantumlike chaos in the frequency distributions of bases A, C, G, T in Human chromosome1 DNA. *APEIRON* Volume 11 Number 3 (July 2004), 134-146. http://redshift.vif.com/JournalFiles/V11NO3PDF/V11N3SEL.PDF. http://arxiv.org/html/physics/0211066

[16] Prigogine, I., and Stengers, I., 1988. *Order Out of Chaos*, third ed. Fontana Paperbacks, London.

[17] Bertalanffy, L. Von, 1968: *General Systems Theory*: Foundations, Development, Applications, George Braziller, New York.

[18] Peacocke, A. R., 1989. *The Physical Chemistry of Biological Organization*, Clarendon Press, Oxford, U. K.

[19] Klir, G. J., 1992. Systems science: a guided tour. J. Biological Systems 1, 27-58.

[20] Berry, M. V., 1988. The geometric phase. *Sci. Amer*. Dec., 26-32.

[21] Maddox, J., 1988: License to slang Copenhagen? *Nature* 332, 581.

[22] Bernaola-Galvan, P., Carpena, P., Roman-Roldan, R., Oliver, J. L., 2002. Study of statistical correlations in DNA sequences. *Gene* 300(1-2), 105-115. http://www.nslij-genetics.org/dnacorr/bernaola02.pdf

[23] Bernardi, G, Olofsson, B, Filipski, J, Zerial, M, Salinas, J, Cuny, G, Meunier-Rotival, M, Rodier F, 1985. The mosaic genome of warm-blooded vertebrates, *Science* 228, 953-958.

[24] Ikemura, T., 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol*. 2, 13-34.

[25] Ikemura, T., Aota, S., 1988. Global variation in G + C content along vertebrate genome DNA: possible correlation with chromosome band structures. *J. Mol. Biol*. 203, 1-13.

[26] Bernardi, G., 1989. The isochore organization of the human genome. *Annu. Rev. Genet*. 23, 637-661.

[27] Ball, P., 2003. Portrait of a molecule. *Nature* 421, 421-422.

[28] Fukushima, A., Ikemura, T., Kinouchi, M., Oshima, T., Kudo, Y., Mori, H., Kanaya, S., 2002. Periodicity in prokaryotic and eukaryotic genomes identified by power spectrum analysis. *Gene* 300, 203-211. http://www.nslij-genetics.org/dnacorr/fukushima02_gene.pdf

[29] Azad, R. K., Subba Rao, J., Li, W., Ramaswamy, R., 2002. Simplifying the mosaic description of DNA sequences. *Physical Review E* 66, 031913 (1 to 6). http://www.nslij-genetics.org/wli/pub/pre02.pdf

[30] Li, W., 2004. A bibliography on 1/*f* noise. http://www.nslij-genetics.org/wli/1fnoise

[31] Li, W., 1997. The study of correlation structure of DNA sequences: a critical review. *Computers Chem*. 21(4), 257-272. http://www.nslij-genetics.org/wli/pub/cc97.pdf

[32] Strachan, T., and Read, A. P., 1996. *Human Molecular Genetics*, βios Scientific Publishers, UK., PP.597.

[33] Li, W., 2002. Are isochore sequences homogeneous? *Gene* 300(1-2), 129–139. http://www.nslij-genetics.org/wli/pub/gene02.pdf

[34] Li, W., Bernaola-Galvan, P., Carpena, P., Oliver, J. L., 2003. Isochores merit the prefix 'iso'. *Computational Biology and Chemistry* 27, 5-10. http://www.nslij-genetics.org/wli/pub/cbc03.pdf

[35] Li, W., and Holste, D., 2004. Universal 1/*f* spectra and diverse correlation structures in Guanine and Cytosine content of Human Chromosomes. Submitted for Journal publication.

[36] Jenkinson, A. F., 1977. *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.

[37] Spiegel, M. R., 1961. Statistics, McGraw-Hill, New York, 359pp.

[38] Selvam, A. M., 1998: 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, **Vol.4**, Singapore, pp.795-809. http://xxx.lanl.gov/abs/chao-dyn/9806001

[39] Cohen, P., 2002. New genetic spanner in the works. *New Scientist* 16 March, 17.

Table 1: Results of power spectral analyses

Set no/ Accession number	Base pairs used for analysis		Number of data sets	Mean C+G concentration per 10bp	Mean T_{50}	Variance spectra following normal distribution (%)	Phase spectra following normal distribution (%)
	from	to					
1 NT_079581.1	1	70000	1	5.47	6.75	100	100
2 NT_079582.1	1	700000	10	4.79	10.20	100	90
3 NT_079583.1	1	560000	8	4.87	8.71	100	100
4 NT_079584.1	1	350000	5	4.61	7.36	100	100
5 NT_011896.8	1	6300000	90	3.82	6.22	96.7	78.9
6 NT_011878.8	1	1050000	15	4.13	7.98	66.7	73.3
7 NT_011875.10	1	9870000	141	3.68	6.34	98.6	83.7
8 NT_011903.10	1	4900000	70	3.95	6.58	95.7	68.6
9 NT_025975.2	1	70000	1	3.89	3.48	100	0
10 NT_079585.1	1	280000	4	3.86	5.94	100	50.0

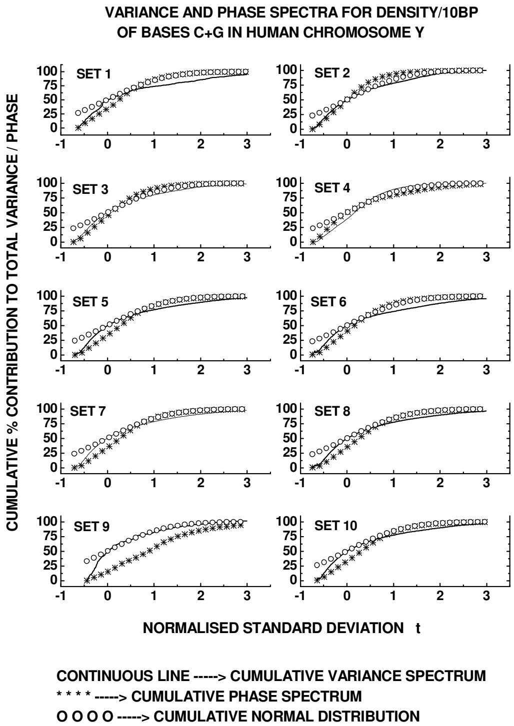

Figure 1: The average variance and phase spectra of frequency distribution of bases C+G in Human chromosome Y for the data sets in each of the 10 contiguous data series given in Table 1. The power spectra were computed as cumulative percentage contribution to total variance versus the normalized standard deviation t equal to ($\log L/\log T_{50}$) –1 where L is the period in years and T_{50} is the period up to which the cumulative percentage contribution to total variance is equal to *50*. The corresponding phase spectra were computed as the cumulative percentage contribution to total rotation (Section 2).

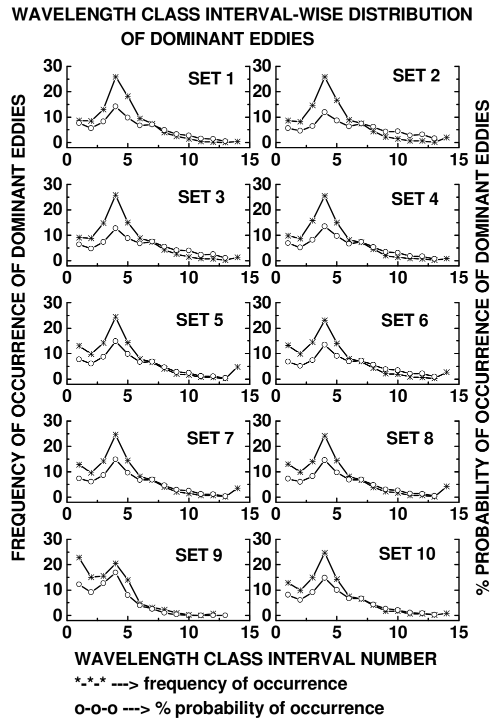

Figure 2: Dominant wavelengths in DNA bases C+G concentration distribution. Average class interval-wise percentage frequency distribution of dominant (normalized variance greater than 1) wavelengths is given by *line + star*. The corresponding computed percentage contribution to the total variance for each class interval is given by *line + open circle*. The observed frequency distribution of dominant eddies closely follow the model predicted computed percentage contribution to total variance