ELSEVIER

BioSystems 49 (1999) 63–70

Bio Systems

Fractal properties of DNA walks

Guillermo Abramson [a,*], Hilda A. Cerdeira [a], Carlo Bruschi [b]

[a] *International Centre for Theoretical Physics, PO Box 586, 34100 Trieste, Italy*
[b] *Interrzational Centre for Genetic Engineering and Biotechnology, Padriciano 99, 34012 Trieste, Italy*

Received 6 October 1997; received in revised form 10 April 1998; accepted 11 May 1998

Abstract

We describe two dimensional DNA walks, and analyze their fractal properties. We show results for the complete genome of *S. cerevisiae*. We find that the mean square deviation of the walks is superdifussive, corresponding to a fractal structure of dimension lower than two. Furthermore, the coding part of the genome seems to have smaller fractal dimension, and longer correlations, than noncoding parts.

Keywords: DNA sequences; Fractals; Random walks; *Saccharomyces cerevisiae*

1. Introduction

There is a growing interest in the scientific community in studying DNA sequences from a physical or mathematical point of view. It has been claimed that long range correlations exist in DNA (Peng et al., 1992; Voss, 1992; Li et al., 1994; Mantegna et al., 1994, 1995), as well as the contrary (Azbel, 1995), and several controversial points have been discussed in the last years. (See, for example, the series of comments that followed Mantegna et al. (1994) in Bonhoeffer et al. (1996), Israeloff et al. (1996), Mantegna et al. (1996) and Voss, (1996).

Long range correlations in DNA sequences were first observed in the two-point autocorrelation function. Moreover, different properties have been found corresponding to the coding and the noncoding portions of the sequences. Usually, correlations are short-ranged and decay exponentially, with one important exception: at the critical point of a phase transition, the exponential turns into a slow-decaying power-law. Many systems evolve, spontaneously, to the critical state. The name 'self-organized criticality' has been coined for them by Bak, whose seminal papers (Bak et al., 1987, 1988) have triggered a major comprehension of how a broad class of such systems

* Corresponding author. Present address: Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Str. 38, 01187 Dresden, Germany

0303-2647/99/$ - see front matter
PII S0303-2647(98)00032-X

work (Bak, 1997). In DNA sequences, not only the existence of this long-ranged correlation is surprising, but also the peculiar form of it. Indeed, it was found to be a $1/f$ spectrum in the Fourier transform of some sequences (Li et al., 1994), that implies a corresponding slow decay in space. This behavior is paradigmatic of self-similar systems, lacking a finite typical length scale.

A large variety of natural systems exhibit long range power-law correlations and scale invariance. Its study provides a good way of quantification of the phenomena: a power law can be quantified with an exponent. The identification of the same exponents in different systems leads to classifications that may otherwise have remained hidden.

As already pointed out by Li et al. (1994), it has to be stressed that a long range statistical correlation does not imply a long-ranged causality in the sequence. It is not the case of a segment of bases affecting some other segment several thousand base pairs away. A long range statistical correlation means that the base density tend to vary in a similar manner in portions of the sequence that lie very far away. Although the biological implications, if any, of these findings are not yet known, we find that the field deserves further investigation. This is in part due to the fact that very long DNA sequences are just now becoming accessible, as a result of the genome projects underway. We focus in this paper on the fractal properties of two dimensional DNA walks, to be defined in Section 2. In particular, most of our results correspond to the organism *S. cerevisiae* (baker's yeast), the first eukaryote whose genome has been completely sequenced (Goffeau et al., 1996). This provides, for the first time, several very long sequences corresponding to the same organism.

2. DNA walk

DNA sequences consist of four symbols, conventionally called A, G, T and C (for adenine, guanine, thymine and cytosine). There is a global bias in the base frequencies. Typically, those of A and T are around 0.30, and those of G and C about 0.20. The frequencies of both A + G (purines) and C + T (pyrimidines) are very close to 0.5. The local density of any of them greatly varies along a sequence, a feature that can be stated as nonstationarity.

There is not a single way of mapping this sequence into a walk (Berthelsen et al., 1992; Mantegna et al., 1995). One of these assigns each symbol to a unit vector in a 4D space. But random walks in high dimensions have the drawback that they are too much 'open' (a true random walk is not compact in a space of dimension higher than two). This is especially bothering since DNA sequences are not arbitrarily long. On the other hand, a mapping into a 1D walk certainly hides some properties of the higher dimensional ones, of which it is a projection. A word of caution against the use of one of the binary sequences has already been raised by Li et al. (1994), which is a review of previous results, including those of two of the authors found in Li and Kaneko (1992). There, it is shown that the fluctuations in density, along the same sequence, of the combined A + G and G + C bases are completely different.

As a compromise, we have chosen a 2D map, built in the following way: each nucleotide is associated with one of these vectors: (1, 0) (0, 1) (− 1, 0) (0, − 1). The walker performs a step in the direction of the vector corresponding to successive nucleotides in the sequence. There are three nontrivial ways of assigning this vectors to the four nucleotides, the others being symmetries of these. They are the following:

$$\begin{aligned}
&\text{(a)}\ T{:}(1,0)\ C{:}(0,1)\ A{:}(-1,0)\ G{:}(0,-1)\\
&\text{(b)}\ T{:}(1,0)\ C{:}(0,1)\ A{:}(0,-1)\ G{:}(-1,0)\\
&\text{(c)}\ T{:}(1,0)\ C{:}(-1,0)\ A{:}(0,-1)\ G{:}(0,1) \qquad (1)
\end{aligned}$$

The bias in base frequency is of course translated into the DNA-walk, and so we have to deal with a biased walk. This bias will always be superimposed to any other feature of the walk, such as correlations, and will surely obscure it. Observe that the mapping marked (a) in Eq. (1) pairs in the same direction the bases A and T, and G and C. The bases of each pair have roughly the same frequency, and to the same extent the map compensates for the bias. Our results are based on

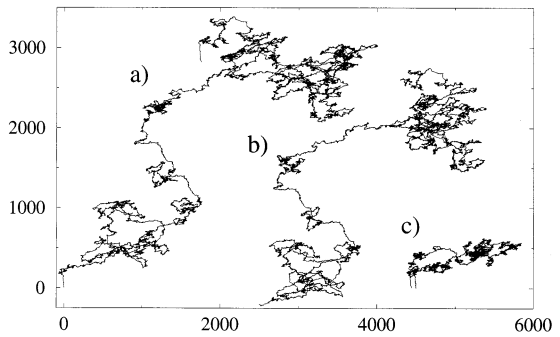

Fig. 1. DNA walk of *S. cerevisiae* chromosome II. The mappings correspond to Eq. (1)a. (a) Whole chromosome. (b) Codenome. (c) Noncodenome. The walks have been shifted for clarity.

the study of this map. A typical DNA walk is shown in Fig. 1a, corresponding to *S. cerevisiae* chromosome II (807188 bases long).

A large portion of the genome of *S. cerevisiae* codes for proteins. Table 1 shows the fraction of coding for each chromosome (about 70%). In higher organisms this quantity drops, down to about 5% in humans (see the gene HUMHBB in Table 1). Using annotated sequences from GenBank (Benson et al., 1998), that can be accessed through the WWW at the National Center for Biotechnology Information (USA) (http://uuu.ncbi.nlm.nih.gov) we have separated each chromosome in two sequences: one consists of the coding fragments (let us call it the 'codenome') and the other of the rest (the 'noncodenome,' formed by intergenic material in general). It has been claimed (Peng et al., 1992) that sequences rich in noncoding material show longer correlations than those composed mostly of coding material. Fig. 1b and c show the DNA walks that correspond to the codenome and the noncodenome of chromosome II. Observe the similarity between the codenome and the whole chromosome (Fig. 1a). Both present filaments that connect 'knots', a feature that gives the figure an extended appearance. On the contrary, the noncodenome appears more compact.

This features will be characterized in what follows. It will be shown that, in *S. cerevisiae*, the

Table 1
Properties of the DNA sequences

Sequence	Length	C	γ	γ_c	γ_{nc}	D	D_c	D_{nc}
S. cerevisiae chromosome I	226646	0.62	1.41	1.47	1.27	1.54	1.50	1.52
S. cerevisiae chromosome II	807188	0.73	1.36	1.40	1.15	1.58	1.60	1.66
S. cerevisiae chromosome III	315341	0.67	1.37	1.40	1.22	1.46	1.45	1.54
S. cerevisiae chromosome IV	1531974	0.73	1.38	1.43	1.16	1.59	1.60	1.63
S. cerevisiae chromosome V	574393	0.67	1.37	1.42	1.12	1.49	1.47	1.65
S. cerevisiae chromosome VI	270149	0.67	1.39	1.43	1.18	1.52	1.50	1.56
S. cerevisiae chromosome VII	1090935	0.72	1.37	1.42	1.11	1.59	1.66	1.64
S. cerevisiae chromosome VIII	562638	0.69	1.37	1.41	1.26	1.52	1.51	1.58
S. cerevisiae chromosome IX	439885	0.72	1.39	1.43	1.14	1.63	1.60	1.62
S. cerevisiae chromosome X	745442	0.75	1.37	1.42	1.11	1.55	1.52	1.70
S. cerevisiae chromosome XI	666448	0.72	1.37	1.42	1.07	1.61	1.55	1.78
S. cerevisiae chromosome XII	1066141	0.70	1.38	1.39	1.31	1.56	1.58	1.53
S. cerevisiae chromosome XIII	924430	0.75	1.37	1.42	1.11	1.60	1.60	1.68
S. cerevisiae chromosome XIV	784328	0.74	1.39	1.43	1.17	1.59	1.60	1.68
S. cerevisiae chromosome XV	1091282	0.72	1.36	1.42	1.14	1.54	1.54	1.59
S. cerevisiae chromosome XVI	948061	0.70	1.36	1.39	1.25	1.69	1.72	1.72
Shuffled chromosome VII	807188	—	1.00	—	—	1.68	—	—
HUMHBB	73308	0.02	1.52	1.40	1.51	1.43	1.51	1.43
Herpes simplex	152260	0.77	1.31	1.28	1.40	1.59	1.60	1.53

The sixteen chromosomes of *S. cerevisiae* are shown, as well as the shuffled chromosome VII, the human β-globin gene and the genome of the *Herpes simplex* virus. The columns display the length of the sequence, the fraction of coding, C, the exponent γ of the mean square fluctuation and the fractal dimension, D, of the whole sequence, the codenome and the noncodenome.

codenome displays the properties of a long-correlated walk, while the noncodenome behaves as a short-correlated one.

3. Mean square deviation

The DNA walks appear very different than ordinary random walks (RW). The 'filaments' that can be seen in Fig. 1 are a hallmark of persistency, a feature not present in uncorrelated RW. The mean square deviation of the walk can be used to quantify this effect. Let's call $r(l)$ the position of the walker at step l. It is a two-dimensional vector since the walk in embedded in a plane. The displacement of the walker, at step l, measured from an arbitrary initial step l_0 is $\Delta r = r(l_0 + l) - r(l_0)$. Its mean square deviation, or fluctuation, is the average of the displacement at step l relative to the mean displacement, that is:

$$F^2(l) = \langle \Delta r^2 \rangle - \langle \Delta r \rangle^2 \quad (2)$$

where the averages are taken over the initial positions l_o. For a wide class of RWs, a power law behavior is expected:

$$F^2(l) \propto l^{\gamma} \quad (3)$$

for some range of values of l (remember that the sequence is finite). An RW without persistence is characterized by $\gamma = 1$, a behavior identified with diffusion. Persistency, or correlations of any range, by $\gamma > 1$. It has already been observed that 1D DNA walks have $\gamma \sim 1.2$ or 1.4 (Peng et al., 1992; Mantegna et al., 1994). In particular, it is claimed in (Peng et al., 1992) that non-coding sequences present a value of γ greater than that of coding sequences, when mapped onto a 1D walk that projects the four bases into purines and pyrimidines.

We have calculated the mean square deviation of the DNA walks derived from whole chromosomes, as well as from the codenome and noncodenome contained in them. The typical behavior can be seen in Fig. 2. The three DNA walks, well above the $F(l)^2 = l$ reference line, clearly show persistency, even though a unique power law behavior is controversial (Abramson et al., 1997). The curve that closely follows the $F(l)^2 = l$ behavior corresponds to an uncorrelated sequence of the same composition as the chromosome. This control sequence is obtained from the DNA one by 'shuffling' its letters a sufficient number of times.

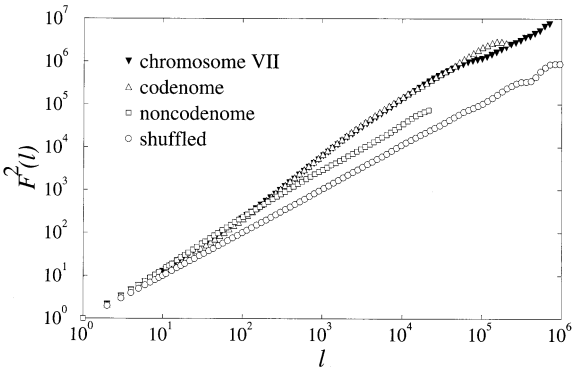

Fig. 2. Mean square deviation of several DNA walks. The sequences correspond to the chromosome VII of *S. cerevisiae*, as indicated in the key. Observe that the codenome is almost identical to the whole chromosome, and that both are above the noncodenome.

It can also be seen that the codenome has almost the same behavior that the whole chromosome, a fact that arises from the high coding content of the yeast genome. The noncodenome, in turn, has its mean square fluctuation below that of the coding, hence showing less persistence, in a more randomlike fashion. The least squares fitted value of γ for each chromosome, and for the corresponding coding and noncoding sequences, are listed in Table 1. The row labeled 'Shuffled chromosome VII' is the control sequence of the chromosome VII. The two last sequences reported in the table are from a human gene (HUMHBB) and a virus (*H. simplex*). In these two cases, the noncoding portion has a greater value of γ than the coding one. This coincides with the behavior reported in (Peng et al., 1992) for the 1D DNA walk. The fact that *S. cerevisiae* has a different behavior, while unexplained, suggests that long range correlations need not be associated with the function of the sequence.

4. Fractal dimension

The fact expressed by Eq. (3) is related to the geometrical structure of the sites visited by the walker (Vicsek, 1989). Consider the set of sites visited during a time t_1, separated at intervals t_2, with $t_1 \gg t_2$. The typical distance from one of these sites to the following one is $l \sim (t_2)^{\gamma}$, since from site to site it makes a random walk. This implies (if the RW of size l do not overlap) that the set can be covered by $N(l)$ boxes of linear size l, so that the number of needed boxes satisfies the scaling relation:

$$N(l) = \frac{t_1}{t_2} \sim l^{-2/\gamma} \tag{4}$$

that is, a fractal dimension $D = 2/\gamma$. If $\gamma = 1$, as in an uncorrelated RW, the fractal dimension is $D = 2$.

The DNA walks are approximate fractal objects. It is not known what could be the biological significance of the self-similarity present in the DNA, but one can speculate that it perhaps arises from a dynamic process that controls its evolution. We have computed the fractal dimension (box-counting and sandbox) of several long DNA walks, including the complete genome of *S. cerevisiae*. The results, given in Table 1, show that this DNA walks have a significantly lower dimension than 2D RW. An RW is compact in two dimensions, with fractal dimension $D = 2$. A finite RW, of course, does not fill the plane, and consequently its measured fractal dimension measured is lower than two. The RW based on shuffled sequences (see the example in Table 1) have a fractal dimension higher than that of the corresponding true DNA walk, though certainly lower than two.

5. Skipped sequences

In an uncorrelated RW, each step is taken independently of the previous steps. In a correlated walk, the direction of each step depends on the history of the previous ones—a feature called 'memory'. The persistency effect displayed by the mean square deviation suggests that DNA walks are of this last kind. Can we build sequences that share some properties of a particular DNA walk but have progressively less memory? We can do this by skipping bases in the original sequence. From the original sequence, we extract subsequences of skip length s by reading the original sequence every s nucleotides. Consider that the original sequence is the series of symbols

$$X_1\ X_2\ X_{3\ldots}X_i$$

Now we define a subsequence of skip length s as X_j, $j = sn + 1$, $n = 0,\ 1,\ 2\ldots$. In this way we generate a set of subsequences:

$s = 1$: $X_1\ X_2\ X_3\ X_4 \ldots X_{n+1}\ldots$

$s = 2$: $X_1\ X_3\ X_5\ X_7 \ldots X_{2n+1}\ldots$

$s = 3$: $X_1\ X_4\ X_7\ X_{10} \ldots X_{3n+1}\ldots$

$s = 4$: $X_1\ X_5\ X_9\ X_{13} \ldots X_{4n+1}\ldots$

s: $X_1\ X_{s+1}\ X_{2s+1}\ X_3 \ldots\ X_{sn+1} \ldots$ with $n = 0,\ 1,\ 2\ldots$

Subsequences of larger skip length share the global features of those with smaller length, but lack their memory up to size s. Since the original sequences (themselves subsequences with $s = 1$) are finite in length, it is clear that the subsequences thus obtained are progressively shorter. This imposes a practical limitation to the accessible values of s, since short sequences give less stable results. But as the subsequences get shorter, we can build more of them from the same original one—there are s different subsequences of skipping length s, starting from s consecutive positions of the original one, and all of them sharing the same global properties. The results that we show in this section are mean values, for each s, taken over these s subsequences.

We have constructed the DNA walks for the skipped subsequences, and measured their fractal dimension. Fig. 3 shows the typical behavior of the fractal dimension for the order s subsequence as a function of the skip length s. The curves are shown only up to $s = 100$ for reasons of clarity. Black squares and full lines correspond to *S. cerevisiae* chromosome VII. Two effects are immediately observed: the fractal dimension grows with growing s, and it does this non-monotonically. The growth of the fractal dimension must be

understood as the effect of losing memory: for growing s, the walk approaches an uncorrelated RW, with a correspondingly higher fractal dimension. Fig. 3 shows also the behavior of the shuffled (uncorrelated) sequence of chromosome VII (open squares), that starts already with a high fractal dimension due to its uncorrelated nature. Both curves, black and white squares, are seen to merge at a skip length near 50. It seems that the persistency does not extend beyond this value.

The other feature, namely the oscillation of period three of the fractal dimension, has its origin in the ternary structure inherent to coding DNA. As mentioned above, the coding content of the yeast genome is rather high, and its structure dominates in this behavior. Although not shown in the figure, the behavior of all the other chromosomes of *S. cerevisiae* share the features exemplified by chromosome VII in Fig. 3.

The codenome of any chromosome (as the one shown in Fig. 3 with circles and dashed lines, corresponding to chromosome VII) has the same oscillations with even greater amplitude, and extending to much larger distances. The noncodenome, on the other hand (triangles, again from chromosome VII), is practically indistinguishable from an uncorrelated RW.

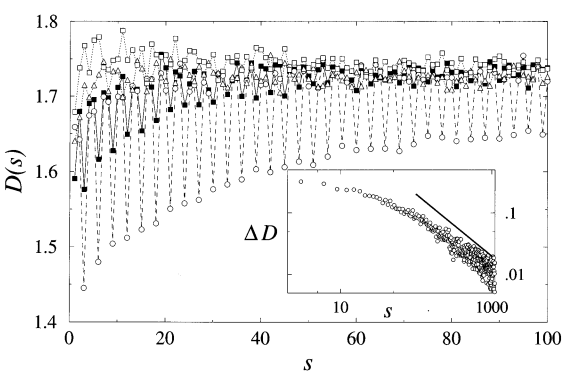

Fig. 3. Fractal dimension of subsequences with skipping length s. The four sets of points correspond to *S. cerevisiae* chromosome VII: the whole chromosome (black squares), the shuffled sequence (white squares), the coding portion (circles) and the noncoding portion (triangles). Inset: log-log plot of the difference between the shuffled and the coding sequences, taken at $s = 371$ (the peaks). A least square fit in the region $100 < s < 2000$ produces an exponent -0.96; a line with slope -1 is shown for reference.

A closer look at the period three, besides, reveals that two relatively high dimensions are followed by a low one. (A similar behavior can be observed in the pair correlation function of any two nucleotides.) This suggests that the coding sequences can be represented as three correlated RW, not correlated between them, and intertwined in such a way that one forms the first nucleotide in every codon, another the second, and the remaining the third. If we represent each one of these by the symbols X, Y and Z, the sequence has the appearance:

$$X_1\ Y_1\ Z_1\ X_2\ Y_2\ Z_2 \ldots X_i\ Y_i\ Z_i$$

In this case, it is easy to see that subsequences with a skipping length $s = 3n$ remain always within the same set. Whereas for s not a multiple of three, the subsequence samples symbols from the three sets. Then, the subsequences where s is a multiple of three will be long correlated and give relatively low fractal dimension—though losing correlation for growing s. Correspondingly, subsequences with s not a multiple of three rapidly lose their correlation due to the mixing of the three uncorrelated sets.

The fractal dimension at the (downward) peaks of Fig. 3, then, characterizes the subsequences with longer correlation. The codenome (circles) has the lowest peaks. What can we say about the range of the correlations? True long range correlations should decay with a slow dependency in space. In contrast, short range—albeit perhaps numerically large, e.g. 1000 bp—are characterized by a fast decay and a typical length, epitomized by an exponential decay. If long range correlations are present, a power law decay could be expected, in contrast to an exponential. As mentioned, the error in the determination of the fractal dimension grows with s due to the shortening of the sequences. This makes it difficult to characterize the functional form of the loss of memory for growing s. Power laws or skewed exponentials fit equally well the lower peaks of Fig. 3, and a decision can not be taken. We have tried another approach, that consists on subtracting the fractal dimension at the peaks of the codenome from the corresponding fractal dimension of the shuffled sequences (in Fig. 3, white squares minus circles),

that can be taken as a 'white noise' reference. This difference is shown in the inset of Fig. 3, in double logarithmic scale. (A line of slope -1, corresponding to the function s^{-1}, is also shown for reference.) Again, we show chromosome VII as an example, but its properties are shared by the whole set of chromosomes of *S. cerevisiae*. It reaches values of s that exceed those shown in the main graph. This log-log plot clearly suggests a power law dependence with an exponent near -1, and rules out an exponential behavior. The long tail of this function implies a long reach of the memory effect in the codenome, and its lack of a finite length scale.

6. Conclusion

We have defined a map of DNA sequences onto 2D pseudo-random walks. We have studied the mean square deviation (MSD) of this DNA walks for the complete genome of *S. cerevisiae*, as well as for some other long DNA sequences. We found that the DNA walks are superdiffusive fractional RW, its MSD growing faster than 1. Moreover, the MSD of the coding part of the sequences grows faster than that corresponding to the non-coding portion. This behavior opposes the one displayed by human sequences, like the, β-globin gene, whose MSD grows faster in the case of the noncodenome.

This feature was further observed at the level of the fractal dimension of the set covered by the DNA walks: the walk that corresponds to coding sequences has lower dimension than the one that comes from the noncoding sequence. The latter is more similar to an uncorrelated RW.

The extraction of subsequences that skip some nucleotides of the original one has shown an increase of the fractal dimension for growing skipping size, that corresponds to the loss of the 'memory' (correlation) effect. This growth is modulated with a period three in coding sequences and in complete sequences with a high coding content, reflecting the underlying codon structure. This loss of memory, finally, is approximately a power-law of exponent -1 when the sequence is compared with the shuffled sequence of the same composition, implying the long range of the correlations and the absence of a finite length scale.

These facts suggest that long range correlations could be associated with the structure of the sequence and not with its function. The origin of the self-similarity in the structure of DNA remains unknown to these authors. But we know that very simple processes can produce the complexity inherent to a scale-free object (e.g. the simple algorithms that produce the textbook fractals (Peitgen et al., 1992). Does some mechanism in the evolution of DNA create the long range correlations? Does the different behavior that we found in the coding and non-coding parts of the complete yeast genome persist in other species? (We have an, extremely, partial answer to this, from the analysis of the human sequence.) How does it relate to the evolutionary distance between species? We believe that these questions will deserve attention in the immediate future, as DNA sequencing continue to provide with more complete material.

Acknowledgements

The authors acknowledge discussions with Pablo Alemany. G.A. also thanks Damian Zanette, Horacio Wio and Ruben Weht for fruitful discussions. H.A.C. acknowledges support of the Istituto Nazionale di Fisica Nucleare (Italy).

References

Abramson, G., Alemany, P.A., Cerdeira, H.A. 1997. Levy-walk analog of two-dimensional DNA-walks for chromosomes of *S. cerevisiae*, (submitted to Phys. Rev. E.).

Azbel, M.Ya., 1995. Universality in a DNA statistical structure, Phys. Rev. Lett. 75, 168–171.

Bak, Per, 1997. How Nature Works. Oxford University Press, Oxford.

Bak, Per, Tang, Chao, Wiesenfeld, K., 1987. Self-organized criticality: an explanation of 1/f noise, Phys. Rev. Lett. 59, 381–384.

Bak, Per, Tang, Chao, Wiesenfeld, K., 1988. Self-organized criticality. Phys. Rev. A 38, 364–374.

Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J., Ouellette, B.F.F., 1998. GenBank, Nuc. Acids Res. 26, 1–7.

Berthelsen, C.L., Glazier, J.A., Skolnick, M.H., 1992. Global fractal dimension of human DNA sequences treated as pseudorandom walks, Phys. Rev. A 45, 8902–8913.

Bonhoeffer, S., 1996. No signs of hidden language in noncoding DNA, Phys. Rev. Lett. 76, 1977.

Goffeau, A., 1996. Life with 6000 genes, Science 274, 546.

Israeloff, N.E., Kagalenko, M., Chan, K., 1996. Can Zipf distinguish language from noise in noncoding DNA?, Phys. Rev. Lett. 76, 1976.

Li, W., Kaneko, K., 1992. Long-range correlation and partial $1/f^{\alpha}$ spectrum in a noncoding DNA sequence, Europhys. Lett. 17, 655–660.

Li, W., Marr, T.G., Kaneko, K., 1994. Understanding long-range correlations in DNA sequences, Physica D 75, 392–416.

Mantegna, R.N., 1994. Linguistic features of noncoding DNA sequences, Phys. Rev. Lett. 73, 3169–3172.

Mantegna, R.N., 1995. Systematic analysis of coding and noncoding DNA sequences using methods of statistical linguistics, Phys. Rev. E 52, 2939–2950.

Mantegna, R.N., 1996. Reply, Phys. Rev. Lett. 76, 1979–1981.

Peitgen, H.-O., Jurgens, H., Saupe, D., 1992. Chaos and Fractals, New Frontiers of Science. Springer-Verlag, New York.

Peng, C.-K., 1992. Long-range correlations in nucleotide sequences, Nature 356, 168.

Vicsek, T., 1989. Fractal Growth Phenomena. World Scientific, Singapore.

Voss, R.F., 1992. Evolution of long-range fractal correlations and $1/f$ noise in DNA base sequences, Phys. Rev. Lett. 68, 3805–3808.

Voss, R.F., 1996. Comment on 'Linguistic features of noncoding DNA sequences', Phys. Rev. Lett. 76, 1978.