ELSEVIER

Physica A 311 (2002) 221–230

PHYSICA A

www.elsevier.com/locate/physa

The fractal structure of the mitochondrial genomes

Nestor N. Oiwa *, James A. Glazier

Department of Physics, Center for the Interdisciplinary Study of Biocomplexity, College of Science, University of Notre Dame, Notre Dame, IN 46556-5670, USA

Received 8 February 2002

Abstract

The mitochondrial DNA genome has a definite multifractal structure. We show that loops, hairpins and inverted palindromes are responsible for this self-similarity. We can thus establish a definite relation between the function of subsequences and their fractal dimension. Intriguingly, protein coding DNAs also exhibit palindromic structures, although they do not appear in the sequence of amino acids. These structures may reflect the stabilization and transcriptional control of DNA or the control of posttranscriptional editing of mRNA.

PACS: ; 87.14.Gg; 61.43.Hv; 64.60.Ak

Keywords: GenBank; Mitochondrial DNA; DNA sequence analysis; Multifractal; Lévy flight

Understanding the relation between biochemical function and the detailed structure of DNA sequences might help us to translate the enormous quantity of sequence data currently being produced into treatments for cancer and viral diseases, new drugs, new genetically engineered crops and animal varieties, and other important practical applications [1]. In this paper, we report that the sequence of the mitochondrial complete genome (mtDNA) presents a well-defined self-similar spatial conformation, associated with typical repeated nucleotide sequences (loops, hairpins and inverted palindromes) in regions coding for functional RNAs and in control regions, including regions relating to the production of transfer ribonucleic acid (tRNA), ribosomal ribonucleic acid (rRNA), and other genes [2]. We also find evidence of self-similarity in

* Corresponding author. Instituto de Fisica, Universidade de Sao Paulo, CP 66318, 05315-970 Sao Paulo, Brazil. Tel.: +55-11-3091-6953; fax: +55-11-3813-4334.

E-mail address: oiwa@if.usp.br (N.N. Oiwa).

0378-4371/02/$ - see front matter
PII: S0378-4371(02)00807-5

mitochondrial and nonmitochondrial protein coding DNA. We interpret GenBank sequences [3] in the form of DNA walks (Lévy flights) using the formalism of diffusive processes and multifractal analysis.

We concentrated our analysis on mtDNA, located in the mitochondria, the organelles responsible for aerobic oxidation in eukaryotes. The complete mtDNA is much shorter than the nuclear DNA, but sufficiently large for statistical analysis for higher multicellular animals. Analysis methods are easier to apply to DNA sequences of this length, around 20,000 base pairs (bp), than to nuclear DNA with its millions of bp. mtDNA has relatively few genes, all of which have been identified, and the biochemical pathways of mitochondria as well as all the proteins resulting from mitochondrial genes are known [2]. mtDNA is a good candidate as biological model to develop our understanding of the relationship between the gene and its environment. We also wish to develop techniques that we can use for phylogenetic analysis. Recent work [4,5] suggests that the structure and organization of DNA sequences relates to the complexity of the phenotype and environmental adaptation [6,7]. Because they are maternally inherited without recombination with paternal DNA, organelle's DNA preserves the evolutionary history of the organism much more clearly than the nuclear DNA [2]. Many studies assume that mutations in nuclear DNA can serve as a rough 'molecular clock' [8]. However, variations occur in the evolution rate of different nuclear genes [9,10]. Thus, phylogentic studies focused on the DNA of organelles (mitochondria or chloroplasts) could be an alternative way to establish the evolutionary timing [4,11]. Finally, the present article does not consider chloroplast DNA (cDNA), because fewer genomes are completely sequenced and we need complete sequences to be able to compare like groups of genes and their control regions.

This background encouraged us to study the relation between the content of mtDNA sequences and the biochemical function of their protein-coding and control sequences. When we talk about 'biochemical expression,' we are supposing that the genome determines phenotypes through biochemicals using a language encoded in the nucleotide sequence. Many approaches have sought this hidden language. The most common approach is to search for similarities between the sequences of genes whose proteins perform the same function in different species or even simply to look for matching sequences without regard to the function of the expressed proteins [12–16]. Histograms counting the appearances of particular nucleotide strings in the genome also allow interspecies comparisons [17,18]. Mantegna et al. applied Zipf's law for the analysis of linguistic texts to noncoding sequences [1] [19].

However, since the early 1990's an alternative sequence analysis technique has applied the theory of diffusive processes to the genome, representing DNA as pseudo-random walks, known as Lévy flights [20–23]. Many types of Lévy flights are possible depending on the different criteria for the steps. The literature usually assigns a right step to purines (cytosine and thymine) and a left step to pyrimidines (adenine and guanine), defining a one-dimensional walk [22]. In the present article, we will use the Berthelsen–Glazier–Skolnick approach [21], because this multidimensional walk permits us to represent the self-similar spatial structures of the DNA double helix,

[1] In their work, the control regions are treated as noncoding.

while the unidimensional walk does not. We construct this walk to obey the following symmetries, based on the biochemical characteristics of the genome: complementarity, reflection, substitution and compatibility [4,21]. When we consider these arguments together in two dimensions ($d=2$), we have right, left, up and down steps for thymine (T), adenine (A), guanine (G) and cytosine (C), respectively. We can construct a Lévy flight in higher embedding dimensions ($d>2$), using similar arguments. Furthermore, we consider just the protein-coding and control regions of the DNA sequence, reducing the interference of 'junk' sequences between genes [5]. Although noncoding regions are few in Metazoan mtDNA, the majority of mtDNA sequences for green algae and plants are neither transcribed into functional RNA, protein coding nor control sequences. We also remark that the embedding space for the DNA walk is not a phase space in the traditional sense of dynamical systems, because we are not describing a movement or oscillation in time, but a discrete space of possible sequences.

For example, we transcribe the *Balaenoptera physalus* mtDNA sequence (finback whale, GenBank accession number NC001321) for valine tRNA (Val-tRNA) in Fig. 1(a) into the walk represented in Fig. 1(b). The histogram in Fig. 2 shows the same walk as Fig. 1(b), but for the complete genome of mtDNA. This mtDNA is 16372 bp long. The height along the z axis is the number of steps inside a box of length 16 bp. All protein-coding (including RNAs) and control segments of the walk are indicated by arrows. Here, we cover the two-dimensional walk using a moving-box algorithm [24]. Instead of a qualitative and subjective description of this irregular object, we can quantify it through its generalized fractal dimensions D_q and their associated Legendre transform, the singularity spectrum $f(\alpha)$.

We estimated D_q and $f(\alpha)$ for the DNA walk using three different methods: box-counting [25–27], moving-boxes [24,28] and sandbox [29]. The first estimates D_q by covering the walk with a set of boxes fixed in a grid. D_q is the slope of the $\ln \ell$ versus $\ln \sum_j p_j^q$ plot, where ℓ is the box size in bp, $p_j = (\#\text{ of stepsin each box})/(\#\text{ of steps of the walk})$ and positive q preferentially weight the dense regions (spikes) of the histogram in Fig. 2, while negative q quantify the sparse regions of the flight. The second algorithm is an improvement on box-counting, which also covers the walk with a set of boxes, but independent of a grid. The boxes can move in space, adapting to the geometry of the analyzed object, reducing errors due to box misalignment. Finally, the sandbox method takes random samples from the walk, and computes D_q through the average slope of $\ln \ell$ versus $\ln \sum_j p_j^q$. Here, ℓ is the diameter in bp of the sample sphere. This method yields results of the same quality as moving-boxes. We estimate $f(\alpha)$ in the same way by fitting the linear portion of a log–log plot [24]. Here, the singularity exponent α weights the light and heavy regions of the walk in the same way as q, but q is an intensive variable, while α is extensive. A simple way to understand $f(\alpha)$ is the following: The density of each occupied point of the walk diverges in a spike at infinitely small radius. Around the point, the density of points in a sphere of radius r decays as $r^{-\alpha}$ so α defines the singularity structure of the spike. $f(\alpha)$ is the fractal dimension of the subset of points of the walk that have singularity α. Bigger α represent the sharp spikes in Fig. 2 while smaller α are the rarefied regions. In a monofractal, all points have the same scaling exponent α and $f(\alpha)$ is just a point with that value of α and f the fractal dimension of the whole object. If the walk were not a

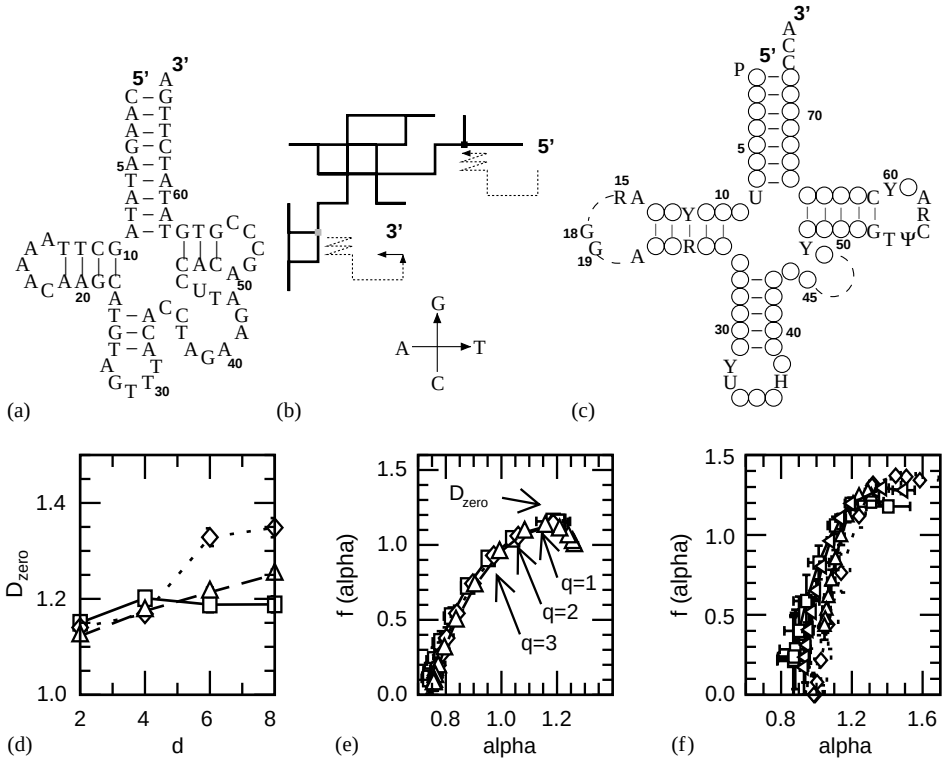

Fig. 1. (a) Suggested conformation for the Val-tRNA of finback whale mtDNA. (b) Two-dimensional walk of (a), where the DNA walks equivalent to the nucleotides 1–9 and 58–66 in (a) are represented by dashed arrows in (b). (c) General diagram for all tRNAs, except for initiator tRNA: A, C, G and U nucleotides (circles), monophosphate group (P), pyrimidine (Y), purine (R), pseudouridine (Ψ), hypermodified purine (H) (extracted from Rich and RajBhandary [35]). (d) D_{zero} for $d = 2, 4, 6, 8$ and (e) $f(\alpha)$ of the two-dimensional Lévy flight of finback whale mtDNA, Fig. 2, where circles, squares and diamonds represent moving-box, box-counting and sandbox algorithms, respectively. (f) $f(\alpha)$ using moving-boxes for nonmitochondrial DNAs: human β-globin region (circles), thale cress chloroplast (squares), coliphage T4 virus (diamonds) and *E. coli* bacteria (triangles).

fractal, our estimate of the slope D_q or $f(\alpha)$ would vary inconsistently with the length scale, and any attempt to fit the slope would result in large errors or unreliable values. However, the three methods give consistent results, Fig. 1(d), and estimate D_{zero} with very small errors for experimental data: $D_{\text{zero}} = 1.14 \pm 0.01, 1.15 \pm 0.01, 1.122 \pm 0.006$ for finback whale mtDNA using box-counting, moving-boxes and sandbox methods, respectively. For most non-DNA experimental multifractals, e.g. in fluid dynamics, we cannot estimate D_q this well and a 10% error is considered a good result which reliably indicates the presence of a multifractal. The error in our DNA D_q estimates is always smaller than 10%, and $< 1\%$ for D_{zero}. The values for D_q at $f(\alpha)$, $1 \leqslant q \leqslant 5$, coincide for each method, Fig. 1(e). However, we cannot confirm the values of $\alpha_{\min}$ obtained for

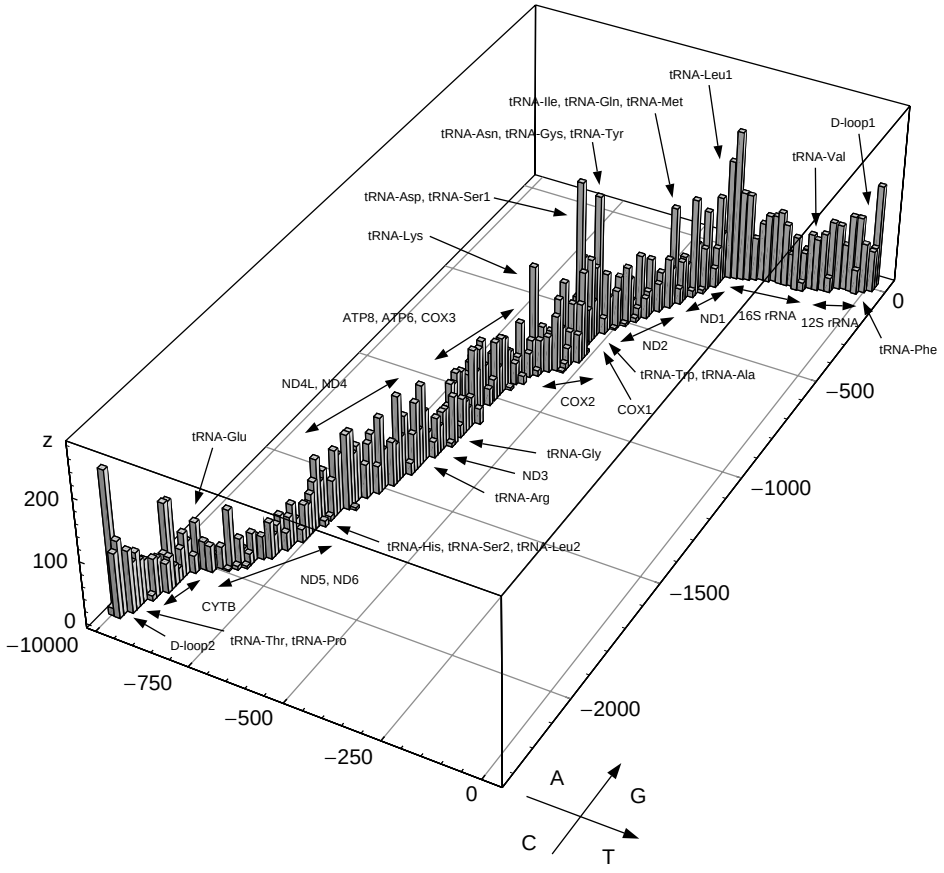

Fig. 2. Two-dimensional walk for the mtDNA of finback whale, where the x-axis represents the DNA walk for guanine (right step) and cytosine (left step), the y-axis, thymine (up step) and adenine (down step), with the number of steps lying inside a box of length 16 base pairs plotted along the z-axis. The arrows indicate the position of each protein-coding (including RNAs) and control segment along the walk.

the sandbox method independently, because the moving-box and box-counting are not reliable in this region. Thus, we have definitively established the multifractal character of the walk.

An initial clue to the structures responsible for the self-similar spatial conformation of mtDNA is the value of D_{zero}. We need to embed in at least two dimensions to represent a fractal of dimension between one and two. If we embedded in one dimension, we could not identify or measure self-similar structures due to the presence of false neighbors [30]. As the Takens theorem suggests, we need to embed in a maximum of 3 or 4 dimensions, since $d < 2D_{\text{zero}} + 1$ [31]. The measured D_{zero} saturates around 1.2 in Fig. 1(b), for embedding dimensions bigger than 2. So, $d = 2$ suffices to describe the walk.

Moreover, we can see from the density plot in Fig. 2 that all tRNAs appear as spikes, with the 12S and 16S rRNAs appearing as the dense structures at the top of the histogram. The secondary interactions between non-neighboring bases due to folding of the DNA (resulting from the two and three hydrogen bonds between A and T and C and G, respectively) of tRNA and rRNA result in a two-dimensional structure (Fig. 1(c)). These structures under tertiary interactions (hydrogen bonds between the arms of tRNA) and the actions of enzymes result in an L-shaped spatial structure with specific biochemical function.

The reason that tRNAs are regions of high density (the nucleotides at positions 1 to 9 and 58 to 66 in Fig. 1(a) and its equivalent in dotted lines Lévy flight in Fig. 1(b)) is that the nucleotide sequence coding for one tRNA arm is followed by its reflected complementary sequence (inverted repeats), and the resulting net DNA walk displacement is null. Usually some nucleotides intervene between the direct and reflected complementary sequences so the walk returns to near its start. In such a situation, a single DNA strand can fold itself into a loop. Each tRNA arm, for example, is a loop. tRNA arms are typically three to eight base pairs long. We find the inverted repeat length n_W in Fig. 1(a) using the following criteria. We start with a long sequence pick a subsequence in it of fixed length, scan through the full sequence and find the nearest occurrence of the inverted subsequence. If the DNA sequence is infinite and non-periodic, we always find the inverted subsequence somewhere, but, the probability of an accidental matching sequence nearby is very low for long subsequences. The probability of an accidental match in the case of a DNA chain composed of four letters increases with the distance i between the given subsequence and the inverted repeat. The probability of an accidental match is $\sum_i (1 - 1/4^{n_W})^i 1/4^{n_W}$ at the ith nucleotide from the original sequence. We accept a maximum 10% chance that any inverted repeat is accidental. So, the probability that the two-dimensional structure with four inverted repeats in Fig. 1(a) is accidental is $< 10^{-4}$.

Furthermore, if we consider the nucleotides (1–12 bp) which do not form part of inverted repeats as a simple two-dimensional random walk, the variance of the associated Lévy flight σ will be 3.8 bp. We can estimate σ from Fig. 1(c). Here, the nucleotides at positions 9, 16–17, 20, 26, 34–36, 38, 44–46, 59 and 73 are not part of inverted repeats. If the DNA sequence were uncorrelated, these 14 bp would contribute a variance of around 3.2 bp, since $\sigma = \sqrt{3n_{\text{steps}}}/2$ in a simple two-dimensional random walk, where n_{steps} is the number of steps of the random walk. Here, we neglect the anticodon which is specific to each individual tRNA, positions from 34 to 36, and use the remaining sequence to construct our walk, since we are not studying specific tRNAs, but the structure common to all tRNAs. The 4 bp at positions 12, 23, 32 and 47 can walk only in two directions, because they must consist only of purines or pyrimidines, adding 2.0 bp more of variance (since $\sigma = \sqrt{n_{\text{steps}}}$ for a one-dimensional walk). Finally, we have some well-defined sequences such as acceptor arms (sequence ACC) that have a specific displacement. Therefore, Fig. 1(c) results in a cluster-like structure for the walk with $\sigma = 3.8$ bp for the entire tRNA, centered at $(0, -1)$ in the space of possible sequences, assuming that the walk started at the origin (0,0).

Despite the remarkable similarity between Figs. 1(a) and (c), Fig. 1(a) is not the real spatial secondary Val-tRNA structure. The walk does not include the unusual bases

present in tRNA (thymine, pseudouridine, dihydrouridine, etc.) or uracil-guanine or adenine-guanine bonds. Although absent from DNA, these bonds are present in tRNA, stabilizing its shape. We also omit the arm due to the sequences GAA and UUC respectively at 39–41 bp and 44–46 bp in Fig. 1(a). Moreover, the acceptor sequence, given by ACC at the end (or beginning) of the sequence, is not present. We do not allow mismatch (or noise) between the two complementary strands, but only exact inverted repeats to reduce the probability of accidental matches. Finally, the anticodon for valine is CAA, and the cruciform diagram shows GUU as the anticodon. This last point indicates that the Val-tRNA is not coding in the direction $5' \to 3'$ of the direct strand, but in the $5' \to 3'$ direction of the complementary strand.

tRNAs are not the only molecules in mtDNA that present loops. rRNAs are also very rich in such highly organized sequences with a huge number of cross-shaped structures, loops and hairpins [2]. Hairpins are loops without nucleotides intervening between the direct and reflected complementary sequences. We can see in Fig. 2 that the walks for 12S and 16S rRNA result in a dense structure at the top of the histogram, reflecting the presence of inverted repeats.

Furthermore, the self-complementary regions defined by loops have an important role in stabilizing circular DNAs. mtDNA is circular like bacterial DNA. In supercoiled double-stranded mtDNA, two large inverted repeat sequences, bigger than 7 bp, form hydrogen-bonded cruciform loops [2]. These two cruciform loops, called inverted palindromes, upon partial denaturation can reduce the linkage number, relaxing the double strand and converting supercoiled DNA to its unsupercoiled relaxed form.

A surprising result is that protein-coding DNA sequences, like nicotine adenine dinucleotide dehydrogenase subunit 1 (gene ND1), also contribute to the self-similar nature of the walk, since they are rich in self-complementary sequences. Fig. 2 shows peaks for ND1, that coincide with inverted nucleotide repeat sequences. We would not expect inverted repeats to contribute to the organization of protein-coding sequences, since the inverted repeats will be lost when the sequence is transcribed into protein, especially since the length of inverted repeats is not in subsets of triplets as it is for codons. We might expect that the final amino acid sequence would be more critical than the intermediate RNA structure for protein coding sequences, so that the palindromes would have to be 'fitted in' where the degeneracy of the DNA codons allows for some sequence flexibility without disturbing the protein sequence (e.g. the degeneracy in the third-base 'wobble' can allow both C–G and A–T rich sequences to code for the same protein. We would not expect this flexibility to be sufficient to allow numerous long tandem repeats, but additional flexibility may come from the substitution of closely related amino acids (hence allowing a much greater effective coding degeneracy) for noncritical sites in a protein (e.g. internal structural sites) [2,32]. However, we found such repeated self-similar structures in nonmitochondrial DNA sequences too. We estimate $f(\alpha)$ using the moving-box method for the human β-globin region (HUMHBB), thale cress chloroplast (AP000423), coliphage T4 virus (AF158101) and *E. coli* bacteria (U00096) in Fig. 1(f). In contrast with Fig. 1(e), where we apply three different methods, we show the $f(\alpha)$ spectrum calculated using just the moving box algorithm, since we have validated the equivalence of the methods in our work with mtDNA. The results for positive qs were confirmed through box-counting (not shown). Unfortunately,

these methods do not allow us to calculate the $\alpha_{\min}$ of $f(\alpha)$ for DNA sequences. In this case, box-counting and moving-box are unreliable for negative q's, and sandbox requires excessive computation times. Furthermore, the sequences for β-globin and coliphage T4 are purely protein coding. Although loops, hairpins, inverted palindromes and other unnamed sequence relations do not play a structural role in the protein, they may play a role in posttranscriptional editing of transcribed raw RNA sequences and in stabilizing and determining the transcription into proteins of the resulting messenger RNA (mRNA). For example, McPheeters et al. predicted stem-and-loop structures in the early lysozyme mRNA of coliphage T4 [33]. While it is difficult to see how such relatively small scale structures could greatly affect the stability and spatial structure of normal double-stranded DNA, they might have a number of effects on the DNA itself, e.g. they might modulate the pitch and writhe of the DNA and affect the histone wrap, thus regulating transcription non-specifically; they might help to determine the local direction of transcription, or they might reflect specific binding sequence motifs overlayed onto the DNA sequence without affecting its coding structure, thus allowing the same DNA sequence to function as both coding sequence and regulator [34]. We will discuss this problem in future work.

We observe identical self-similarity in mtDNA of algae and plants: *Chlamydomonas eugametos* (green algae, AF008237), *Pedinomonas minor* (green algae, AF116775), *Marchantia polymorpha* (liverwort, M68929), and *Arabidopsis thaliana* (thale cress, Y08502). Fungi, as *Hansenula wingei* (D31785), *Podospora anserina* (X55026), *Schizosaccharomyces pombe* (X54421), and *Saccharomyces cerevisiae* (baker's yeast, NC001224), also present a well-defined $f(\alpha)$ curve. We found the same behavior in nematodes (worms) too: *Caenorhabditis elegans* (NC001328), *Ascaris suum* (pig roundworm, NC001327), and *Onchocerca volvulus* (river blindness roundworm, NC001861). We could track the self-similarity of insects, *Drosophila yakuba* (fly, NC001322), *Drosophila melanogaster* (fruit fly, NC001709), *Ceratitis capitata* (mediterranean fruit fly, NC000857), and *Apis mellifera* (honey bee, NC001566), computing the fractal dimension for the annelid, *Lumbricus terrestris* (common earthworm, NC001673). However, the best multifractal tracking is for the chordate branch of the Metazoan phylogenic tree. This branch starts with *Strongylocentrotus purpuratus* (purple sea urchin, NC001453), and *Paracentrotus lividus* (common urchin, NC001572). Then, we must consider the Hemichordate *Balanoglossus carnosus* (acorn worm, NC001887), followed by the chordate *Branchiostoma lanceolatum* (amphioxus, NC001912). After that, we have lampreys as *Petromyzon marinus* (sea lamprey, NC001626). Sharks are the next evolutionary step and the mtDNA Lévy flight of *Mustelus manazo* (gummy shark, NC000890) is a multifractal object. The fishes, *Cyprinus carpio* (common carp, NC001606), and *Crossostoma lacustre* (tasseled-mouth loach, NC001727), as well as the amphibian *Xenopus laevis* (African clawed frog, NC001573) are also self-similar. The DNA walk of American alligator (*Alligator mississippiensis,* NC001922), chicken (*Gallus gallus,* NC001323), and the monotremata *Ornithorhynchus anatinus* (duckbill platypus, NC000891) are multifractal too. For mammals, we do not limit our computation to finback whale, but we extend the analysis to more six species: *Mus musculus* (mouse, NC001569), *Rattus norvegicus* (Norway rat, NC001665), *Bos taurus* (cow, NC001567), *Phoca vitulina* (harbor seal, NC001325), *Homo sapiens* (human,

NC001807), and *Balaenoptera musculus* (blue whale, NC001601). We thus see true multifractality in all 35 mtDNAs analyzed showing that self-similarity is independent of level of evolutionary complexity.

Loop-, hairpin- and inverted palindrome-rich structures, like tRNA and rRNA, as well as protein-coding and control sequences which also contain inverted repeats, are responsible for the cluster-like structures we observe in DNA sequence Lévy flights. We can characterize the clustering of the DNA walk using multifractal analysis. The multifractality of mtDNA and the $f(\alpha)$ spectrum for nonmitochondrial DNA suggest a general multiscale nonlinear organization for both coding and control sequences. The biological mechanism which generates this structure, its function and the relation between codon choice in protein-coding sequences and the inverted repeated relations are objects of our current research.

This work was supported by: Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil, under grants 1998/05253-2, 1999/08479-4, 2001/05839-1 and 2001/13390-4; National Science Foundation, Division of Integrative Biology, The United States of America, IBN-0083653; NASA Glenn Research Center, USA, NAG3-2366; Department of Energy, USA, DE-FGO299ER45785; National Science Foundation & Conselho Nacional de Desenvolvimento Cientıifico e Tecnológico International Award, USA-Brazil, INT98-02417.

References

[1] E.S. Lander, et al., Nature 409 (2001) 860.
[2] J.D. Watson, N.H. Hopkins, J.W. Roberts, J.A. Steiz, A.M. Weiner, Molecular Biology of The Gene, 4th Edition, The Benjamin/Cummings, Menlo Park, CA, 1987.
[3] D.A. Benson, I.K. Mizrachi, D.J. Lipman, J. Ostell, B.A. Rapp, D.L. Weeler, Nucleic Acids Res. 28 (2000) 15–18.
[4] J.A. Glazier, S. Raghavachari, C.L. Berthelsen, M.H. Skolnick, Phys. Rev. E 51 (1995) 2665–2668.
[5] N.N. Oiwa, C. Goldman, Phys. Rev. Lett. 85 (2000) 2396–2399.
[6] G. Nicolis, I. Prigogine, Exploring Complexity, Freeman, New York, 1989.
[7] H. Haken, Information and Self-Organization—A Macroscopic Approach to Complex Systems, Springer, Berlin, 1988.
[8] F.J. Ayala, A. Rzhetsky, F.J. Ayala, Proc. Natl. Acad. Sci. 95 (1998) 606–611.
[9] F.J. Ayala, Proc. Natl. Acad. Sci. 94 (1997) 7776–7783.
[10] F. Rodriguez-Trelles, Rosa Tarrio, F.J. Ayala, Proc. Natl. Acad. Sci. 98 (2001) 11405–11410.
[11] M.W. Gray, D. Sankoff, R.J. Cedergren, Nucleic Acids Res. 12 (1984) 5837.
[12] R. de Rosa, J.K. Grenier, T. Andreeva, C.E. Cook, A. Adoutte, M. Akam, S.B. Carroll, G. Balavoine, Nature 399 (1999) 772.
[13] M.V. Penkina, O.I. Karpova, S.Ya. Dadashev, N.V. Mil'shina, Yu.F. Bogdanov, Gene Mol. Biol. 31 (1997) 198.
[14] J. Nieto-Sotelo, K.B. Kannan, L.M. Martıinez, C. Segal, Gene 230 (1999) 187.
[15] H.-S. Lee, J.-H. Mun, S.-G. Kim, Gene 226 (1999) 155.
[16] S.J. Russek, Gene 227 (1999) 213.
[17] B.-L. Hao, H.C. Lee, S.-Y. Zhang, Chaos Solitons Fractals 11 (2000) 825;
Z.-G. Yu, B.-L. Hao, H.-M. Xie, G.-Y. Chen, Chaos Solitons Fractals 11 (2000) 2215.
[18] M. Kolwalczuk, A. Gierlik, P. Mackiewicz, S. Cebrat, M.R. Dudek, Physica A 273 (1999) 116.
[19] R.N. Mantegna, S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, M. Simons, H.E. Stanley, Phys. Rev. Lett. 73 (1994) 3169.
[20] M.A. Gates, J. Theoret. Biol. 119 (1986) 319–328.

[21] C.L. Berthelsen, J.A. Glazier, M.H. Skolnick, Phys. Rev. A 45 (1992) 8902–8913.
[22] S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, M. Simons, H.E. Stanley, Phys. Rev. E 47 (1993) 4514–4523.
[23] A. Arnéodo, Y. d'Aubenton-Carafa, E. Bacry, P.V. Graves, J.F. Muzy, C. Thermes, Wavelet based fractal analysis of DNA sequences, Physica D 96 (1996) 291.
[24] N.N. Oiwa, N. Fiedler-Ferrara, Physica D 124 (1998) 210–224;
N.N. Oiwa, N. Fiedler-Ferrara, Phys. Rev. E 65 (2002) 036702.
[25] A. Block, W. von Bloh, H.J. Schellnhuber, Phys. Rev. A 42 (1990) 1869–1874.
[26] X.-J. Hou, R. Gilmore, G.B. Mindlin, H.G. Solari, Phys. Lett. A 151 (1990) 43–46.
[27] L.V. Meisel, M. Johnson, P.J. Cote, Phys. Rev. A 45 (1992) 6989–6996.
[28] M. Yamaguti, C.P.C. Prado, Phys. Rev. E 55 (1997) 7726–7732.
[29] T. Tél, A. Fülöp, T. Vicsek, Physica A 159 (1989) 155–166.
[30] M.B. Kennel, R. Brown, H.D.I. Abarbanel, Phys. Rev. A 45 (1992) 3403–3411.
[31] F. Takens, Lecture Notes Math. 898 (1980) 366–381.
[32] J.E.M. Hornos, Y.M.M. Hornos, Phys. Rev. Lett. 71 (1993) 4401–4404.
[33] D.S. McPheeters, A. Christensen, E.T. Young, G. Stormo, L. Gold, Nucleic Acids Res. 14 (1986) 5813–5826.
[34] M. Takahashi, J. Theoret. Biol. 141 (1981) 117–136.
[35] A. Rich, U.L. RajBhandary, Ann. Rev. Biochem. 45 (1976) 805–860.