DNA in chromatin: how to extract structural, dynamical and functional information from the analysis of genomic sequences using space-scale wavelet techniques

Alain Arneodo

Laboratoire de Physique, Ecole Normale Supérieure de Lyon
46 allée d'Italie, 69364 Lyon Cedex 07, FRANCE

Françoise Argoul
Benjamin Audit
Samuel Nicolay
Edward-Benedict Brodie

Cédric Vaillant
EPF Lausanne, Switzerland

Marie Touchon
Yves d’Aubenton-Carafa
CGM, Gif-sur-Yvette, France

Claude Thermes
DNA in chromatin: how to extract structural, dynamical and functional information from the analysis of genomic sequences using space-scale wavelet techniques

Laboratoire de Physique, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, FRANCE

Approved for public release, distribution unlimited

See also ADM001750, Wavelets and Multifractal Analysis (WAMA) Workshop held on 19-31 July 2004.

The original document contains color images.
DESOXYRIBONUCLEIC ACID
A FEW HISTORICAL LANDMARKS

1869 Miescher isolates DNA

1944 DNA carries the genetic information (Avery)

1953 The double helix structure of DNA is discovered by Watson and Crick

→ a simple model for the transmission of the genetic information

1966 Niremberg, Ochoa and Khorana elucidate the genetic code

→ DNA codes for proteins

<table>
<thead>
<tr>
<th>codon</th>
<th>ATG</th>
<th>GCG</th>
<th>ACG</th>
<th>...</th>
<th>GCC</th>
<th>GTG</th>
<th>TAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>amino acid</td>
<td>Met</td>
<td>Ala</td>
<td>Thr</td>
<td>...</td>
<td>Ala</td>
<td>Val</td>
<td>stop</td>
</tr>
</tbody>
</table>
DeoxyriboNucleic Acid

DNA Molecule: Two Views

- Double helix macromolecule
- Each strand consists of an oriented sequence of four possible nucleotides: Adenine, Thymine, Guanine & Cytosine
- Complementary strands: \([A]=[T]\) & \([G]=[C]\) over the sum of both strands
Organization of the Human Genome

Transcription

Maturation

Traduction

23 Chromosomes
$L \sim 100$Mbp.

Genes ($\sim 20\%$)
$L \sim 10$kbp.

Introns
(INTervening seq.)
$L \sim 1$kbp.

Exons
(EXpressed seq.)
$L \sim 150$bp.

Proteins
$L \sim 500$AA.

Non genic DNA
Sequencing projects result in 4 letter texts:

gtcagtttctgaggcgggtcgggacccaggcgtgagctggactgcctgcaccggcggccagtgcgttcctcctctcgactgcgtcttggtggcgccagga
cgcggccggggtgcggccggcgcctttcctggagatgggtgcgcaccactcctcc
tcgagaagccagggagacagcccagacccccgacagccttcaggctctggccagtcc
gtcagattccccctagggtccagggaggaccccgctaatccagctccctctc
gggggtactggggaccccttcgagctccactccccatcagggtccagggga
gaccccccaactatgtcctaggggtccaggggagatggccagccaccccccaact
cgcgtttccctggggcccccccctccccctttacagctcaaccttcctccctcgagagt
cggggggctgggggtctcgttcagttcctgagttccccctccccccctccctccgggtgtcc
cggggggcgcggccccccgaccccccaactgtctgtgtattcccccaaggcgccgggtct
cggccgcagctgtctccacgctctgtctctctgctctctcttcgcctcagtcc
gttctatctttttcatcttttatattgactctgcgcccaaatattctttgact
tctatttagagaaggattcgggggtctatattttattcgggggcgtgtgcttat
aagtccaaacagatgagattttcagttataatccttcgggtgacttatgctattaatgcaccataagccagaaataggacttcagtgccgaaaactg
cattccggctctgtctctgtctgtctctgtccaccccttggcaatgtttcgggcgg
cgggaagcgctgcactcgccgacgcgtgacgtaacctgtttcttgtcttccagg
tcatacgccctctgtggggtggcgcagattgacatataactttttactattttttata
ctggtaagcaacnlacaattttcttcagcagggggtctgtgataaaaatagara
tgcctactctcggagctctgaagccatccttaagcctccccctgtatgtgccc
ccatgtcatctctccccctcatacgggctcttgactgtagctgacacctgtgtgataaa
actggtataatagcaatgaactttttttttttgtgctctgtaagctctcttag
caaattactctaggaggaggggtctttgggaaaccctctgattatataaggccggg
cagtcagcagcactacagctgccgcaagattgttggttcctgaagagcagcatccagaggtgg
gcagtagggtgggacgtgagcccttcacaccttgggggtctggccctgtcaag
gcagtgctcagaatgtaagtgaatgttggacgggtcggtgtcagagacttg
gagaacttgtggtttggtgtgtaaaaaactnacatatttagggctcagaagtag...
NET RESULT: EACH DNA MOLECULE HAS BEEN PACKAGED INTO A MITOTIC CHROMOSOME THAT IS 50,000x SHORTER THAN ITS EXTENDED LENGTH
DIFFERENT WAYS TO READ THE TEXT

I. “Classical” reading

- Looking for patterns
  - Genes, introns, exons detection
  - Splicing sites, promoters, replication origins recognition

- Characterizing repetitions
  - Tandem, interspersed repeats
  - Oligonucleotide usage

- Using methods such as
  - Hidden Markov chains
  - Fourier transform
  - Dot-plot matrices and recurrence plots

INvariance UNDER TRANSLATION
II. The physicist reading

- **Hypothesis:** The DNA text results from a stochastic process:
  
  ACGTTTCGAT?

- **Question:** The choice of the next nucleotide:
  
  i. Depends on a **finite** number \( l_o \) of the previous trials
     
     → **Short range** correlations and **exponential** decay of the correlation function:
     
     \[
     C(l) \propto \exp(-l/l_o)
     \]

  ii. Depends on all the previous nucleotides
     
     → **Long range** correlations and **power law** decay of the correlation function:
     
     \[
     C(l) \propto l^{-\kappa}
     \]

INVARINANCE UNDER DILATATION
DNA WALK REPRESENTATION (PENG et al. 92)

1. Each nucleotide is associated to a numerical value (A to a, T to t, G to g and C to c).

   purine-pyrimidine : \(a = g = 1\) and \(t = c = -1\)
   weak-strong : \(a = t = 1\) and \(g = c = -1\)
   amino-keto : \(a = c = 1\) and \(t = g = -1\)

   A-non A : \(a = 1\) and \(t = g = c = -1/3\)
   T-non T : \(t = 1\) and \(a = g = c = -1/3\)
   G-non G : \(g = 1\) and \(a = t = c = -1/3\)
   C-non C : \(c = 1\) and \(a = t = g = -1/3\)

2. Suppose you have a walker on the line. The value associated to the \(i^{th}\) nucleotide defines the \(i^{th}\) step \(S(i)\) of the walker.

Example using the purine (↑) pyrimidine (↓) distinction:

\[ f(n) = \sum_{i=1}^{n} S(i) \]
Most of the physicist works amount to characterizing the roughness of a DNA walk landscape
Most of the physicist works amount to characterizing the roughness of a DNA walk landscape.
FRACTAL SIGNALS

- Turbulent velocity signal: $V(t)$
- Brownian signal: $S(t)$ ("1/f noise")
- Medical signal: Heart rate
- Financial time series: Market prices
ROUGHNESS EXPONENT

- Root-mean square of the height fluctuations:
  \[ W(L) = \sqrt{< f^2(x) > - < f(x) >^2} \sim L^H \]
  \[ H = \text{roughness exponent} \]
  \[ D_f = 2 - H \]

- Random walk
  - \(0.5 < H < 1\) \ LONG RANGE CORRELATIONS (LRC)
  - \(H = 0.5\) \ UNCORRELATED
  - \(0 < H < 0.5\) \ ANTI-CORRELATIONS

- Power spectrum
  \[ S_f(k) \sim k^{-(2H+1)} \]

- Correlation function
  \[ C_f(l) = < f(x) f(x+l) > - < f(x) >^2 \sim l^{2H} \]
Are the observed LRC a bias in the measurement?

Is the mosaic structure of DNA enough to account for the observed misleading LRC in DNA sequences?

Karlin and Brendel 93:

A specific analysing tool is needed to avoid confusing a biased uncorrelated random walk with an unbiased correlated random walk.
WAVELET ANALYSIS OF FRACTAL SIGNALS

\[ T_{g}(a,b) = \frac{1}{a} \int g^{*} \left( \frac{x-b}{a} \right) f(x) \, dx \]

Mathematical microscope

“Singularity scanner”

The wavelet transform allows us to LOCATE (b) the singularities of f and to ESTIMATE (a) their strength \( h(x) \) (Hölder exponent)
CONTINUOUS WAVELET TRANSFORM OF THE TRIADIC DEVIL’S STAIRCASE

The Devil’s Staircase

\[ F(x) = \int_{-\infty}^{x} d\mu(x) \]

Wavelet Transform Modulus Maxima (WTMM)

WTMM Skeleton

\[ F(x) \text{ is continuous but non differentiable. } F'(x) = 0 \text{ almost everywhere. Its continuous variation occurs over a set of Lebesgue measure } = 0 \text{ and dimension } D_F = \log 2 / \log 3 \]
**Fractal measures**

- Invariant measures associated with the strange attractors of discrete dynamical systems
- Turbulent energy dissipation

**TRIADIC CANTOR SET**

- **F(x)** is continuous but non-differentiable. \( F'(x) = 0 \) almost everywhere.
- Its continuous variation occurs over a set of Lebesgue measure = 0 and dimension \( D_F = \log 2 / \log 3 \)

**Fractal signals**

- Weierstrass functions
- Fractional Brownian motions
- Turbulent signals

**DEVIL’S STAIRCASE**

\[ F(x) = \int_{-\infty}^{x} d\mu(x) \]

Characteristic function of \( \mu \)

\[ F(x) = \int_{-\infty}^{x} d\mu(x) \]
WAVELET ANALYSIS OF THE DNA SEQUENCE OF THE BACTERIOPHAGE $\lambda$
SYNTHETIC DNA SEQUENCES

Uncorrelated random sequence

Long range correlated random sequence

Purine

$w = 32\text{bp}$

Purine excess

$w = 32\text{bp}$

Purine

$w = 512\text{bp}$

Purine excess

$w = 512\text{bp}$
SYNTHETIC DNA WALKS

Fractional Brownian motions: $B_H$

- $H = 0.3$ anti-correlated
- $H = 0.5$ uncorrelated
- $H = 0.7$ long-range correlated
- $H = 0.9$ long-range correlated
A UNIQUE WAY TO DISPLAY RESULTS

1. Straight line $\Leftrightarrow$ scale invariance properties

2. The slope of a linear behavior gives the roughness exponent $H$

$$\begin{align*}
H &= 0.5 \quad \text{No LRC} \\
H &> 0.5 \quad \text{LRC}
\end{align*}$$
A UNIQUE WAY TO DISPLAY RESULTS

1. Straight line $\Leftrightarrow$ scale invariance properties

2. The slope of a linear behavior gives the roughness exponent $H$

\[
\begin{cases} 
H = 0.5 & \text{No LRC} \\
H > 0.5 & \text{LRC}
\end{cases}
\]
LRC and the Isochore Structure of Warm Blooded Vertebrates

G + C poor  G + C rich

Introns  Exons  3rd bases

Human  Rodents  Birds  Drosophila

LRC increase with the $G + C$ content of isochores
This result remains valid for genomes that don’t possess an isochore structure!
WHICH BIOLOGICAL MECHANISMS CAN ACCOUNT FOR LRC IN DNA SEQUENCES

- Genomes dynamics and plasticity
  - Point mutation
  - Insertion, deletion
  - Transposition
  - Duplication of exons, genes or chromosomes
  - Recombinaison
    - Generalized Lévy walk model (Buldyrev et al. 93)
    - Length distribution of protein coding segments (Herzel and Große 97)

- Compaction constraints - Accession to information
  - Nucleosome
  - Chromatine fiber
  - Higher order folding up to the metaphase chromosome
    - Fractal model of chromosomes (Takahashi 89)
    - Crumpled globule model (Grosberg et al. 93)
HIERARCHICAL STRUCTURE OF EUKARYOTIC DNA

- Short region of DNA double helix: 2 nm
- "beads-on-a-string" form of chromatin: 11 nm
- 30-nm chromatin fiber of packed nucleosomes: 30 nm
- Section of chromosome in an extended form: 300 nm
- Condensed section of metaphase chromosome: 700 nm
- Entire metaphase chromosome: 1400 nm
Statistical analysis of the eukaryotic genome of Saccharomyces cerevisiae

Universality between the 16 chromosomes of yeast
Universality between the 4 mononucleotidic codings
\( n_c \sim 200\text{bp} \) is a characteristic length scale
Yeast chromosome I

Gaussian statistics at small scales \((n \leq 200 \text{bp})\)

Non-Gaussian (fat tails) statistics at large scale \((n \geq 200 \text{bp})\)
**STATISTICAL ANALYSIS OF THE BACTERIAL GENOME OF Escherichia coli**

Universality between the 4 mononucleotidic codings and with the eukaryotic genome of yeast

\( n_c \sim 200\text{bp} \) is a characteristic length scale

Gaussian statistics at small scales \( (n \leq 200\text{bp}) \): \( H = 0.5 \)

Non Gaussian (fat tails) statistics at large scale \( (n \geq 200\text{bp}) \): \( H = 0.75 \)
DNA walks that reflect the structure of the DNA polymer

2 trinucleotide codings based on experiments:

<table>
<thead>
<tr>
<th>Trinucleotide</th>
<th>PNum</th>
<th>DNase I</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA/TTT</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>AAC/GTT</td>
<td>3.7</td>
<td>1.6</td>
</tr>
<tr>
<td>AAG/CTT</td>
<td>5.2</td>
<td>4.2</td>
</tr>
<tr>
<td>AAT/ATT</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>ACA/TGT</td>
<td>5.2</td>
<td>5.8</td>
</tr>
<tr>
<td>ACC/GGT</td>
<td>5.4</td>
<td>5.2</td>
</tr>
<tr>
<td>ACG/CGT</td>
<td>5.4</td>
<td>5.2</td>
</tr>
<tr>
<td>ACT/AGT</td>
<td>5.8</td>
<td>2.0</td>
</tr>
<tr>
<td>AGA/TCT</td>
<td>3.3</td>
<td>6.5</td>
</tr>
<tr>
<td>AGC/GCT</td>
<td>7.5</td>
<td>6.3</td>
</tr>
<tr>
<td>AGG/CCT</td>
<td>5.4</td>
<td>4.7</td>
</tr>
<tr>
<td>ATA/TAT</td>
<td>2.8</td>
<td>9.7</td>
</tr>
<tr>
<td>ATC/GAT</td>
<td>5.3</td>
<td>3.6</td>
</tr>
<tr>
<td>ATG/CAT</td>
<td>6.7</td>
<td>8.7</td>
</tr>
<tr>
<td>CAA/TTG</td>
<td>3.3</td>
<td>6.2</td>
</tr>
<tr>
<td>CAC/GTG</td>
<td>6.5</td>
<td>6.8</td>
</tr>
<tr>
<td>CAG/CTG</td>
<td>4.2</td>
<td>9.6</td>
</tr>
<tr>
<td>CCA/TGG</td>
<td>5.4</td>
<td>0.7</td>
</tr>
<tr>
<td>CCC/GGG</td>
<td>6.0</td>
<td>5.7</td>
</tr>
<tr>
<td>CGG/CGG</td>
<td>4.7</td>
<td>3.0</td>
</tr>
<tr>
<td>CGA/TCG</td>
<td>8.3</td>
<td>5.8</td>
</tr>
<tr>
<td>CGC/GCG</td>
<td>7.5</td>
<td>4.3</td>
</tr>
<tr>
<td>CTA/TCG</td>
<td>2.2</td>
<td>7.8</td>
</tr>
<tr>
<td>CTC/GAG</td>
<td>5.4</td>
<td>6.6</td>
</tr>
<tr>
<td>GAA/TTC</td>
<td>3.0</td>
<td>5.1</td>
</tr>
<tr>
<td>GAC/GTC</td>
<td>5.4</td>
<td>5.6</td>
</tr>
<tr>
<td>GCA/TGC</td>
<td>6.0</td>
<td>7.5</td>
</tr>
<tr>
<td>GCC/GGC</td>
<td>10.0</td>
<td>8.2</td>
</tr>
<tr>
<td>GGA/TCC</td>
<td>3.8</td>
<td>6.2</td>
</tr>
<tr>
<td>GTA/TAC</td>
<td>3.7</td>
<td>6.4</td>
</tr>
<tr>
<td>TAA/TTA</td>
<td>2.0</td>
<td>7.3</td>
</tr>
<tr>
<td>TCA/TGA</td>
<td>5.4</td>
<td>10.0</td>
</tr>
</tbody>
</table>
1. Nucleosome positioning model (PNuc)

related to curvature?

2. DNase I digestion data related to bending propensity

Hypothesis: LRC in the small scales regime is the signature of the nucleosomal structure
Eucaryotes  Bacteria

Human  Haemophilus influenzae

Drosophila melanogaster  Treponema pallidum

Arabidopsis thaliana  Bacillus subtilis

$h(q,n) - 0.6 \log_{10} n$

(-- --) DNA text, (o) PNUC, (■) DNase I

Nucleosomes  No nucleosomes
Small scales LRC are related to nucleosome-like structures.

Pox virus don’t display LRC in the small scale regime.

Archaebacteria display LRC in the small scale regime.
AFM visualisation of a reconstituted chromatin fiber

Pierre-Louis Porté, Emeline Fontaine, Cendrine Moskalenko

Images obtained in ‘Tapping Mode’ in air
Linear DNA (2500 bp) positioning nucleosomes

Image obtained in ‘Tapping Mode’ in air
Linear DNA (2500 bp) positioning nucleosomes

Image obtained in ‘Tapping Mode’ in air
Plasmid DNA (3200 bp) + nucleosomes

Images obtained in ‘Tapping Mode’ in air
Plasmid DNA (3200 bp) + nucleosomes

Images obtained in ‘Tapping Mode’ in air
1. Nucleosome positioning model (PNuc)

related to curvature?

2. DNase I digestion data
related to bending propensity

Hypothesis: LRC in the small scales regime is the signature of the nucleosomal structure
HIERARCHICAL STRUCTURE OF EUKARYOTIC DNA

- Short region of DNA double helix
- "beads-on-a-string" form of chromatin
- 30-nm chromatin fiber of packed nucleosomes
- Section of chromosome in an extended form
- Condensed section of metaphase chromosome
- Entire metaphase chromosome
Large Scale Representation of Genomic Sequences

Space-Scale Representation of the GC Content with a Smoothing Gaussian Filter

Chromosome 22 (Human)

Filtering scales: $a_1^* = 40$kb, $a_2^* = 160$kb

Space-scale content: $S(a) = \sum_n |T_{\psi_M}(n, a)|$, where $\psi_M$ is the Morlet wavelet
Transcription

Opening of the double helix with a different environment for each strand => asymmetrical process

Replication

Semi-conservative Replication

Opening of the double helix with a different environment for each strand => asymmetrical process
Symmetrical properties of the strands:
“Parity Rule type 2”

\[ [A] = [T] \quad \& \quad [G] = [C] \]

\textit{in each strand}

Deviations from this property estimated by the compositional skews

\[ S_{CG} = \frac{[C] - [G]}{[C] + [G]} \]

\[ S_{AT} = \frac{[A] - [T]}{[A] + [T]} \]

Compositional skew due to local biases in a strand in the course of biological mechanisms
Strand Compositional Asymmetry

Chromosome 22 (Human)

\[ T_{90}(n, a^*_2) \]

- sense genes
- anti-sense genes
- non-coding sequences

Filtering scales: \( a_1^* = 40\text{kb}, a_2^* = 160\text{kb} \)
A wavelet based methodology to detect gene clusters

Chromosome 22 (Human)

Analyzing wavelet: \[ g^{(n)}(x) = \frac{1}{\sqrt{2\pi}} \frac{d^n e^{-x^2/2}}{dx^n} \]

\[ T_{g^{(n)}}(b, a) = \frac{1}{a} \int f(x) g^{(n)}(\frac{x-b}{a}) \, dx = \frac{d^m}{db^n} T_{g^{(0)}}(b, a) \]
A wavelet based methodology to detect replication origins

Experimentally observed replication origin in the human genome

**Globin:** 4008 kb  
**Chromosome 11**  
**Predicted RO:** 4009 kb

![Graph showing skew and scale (bp) for A/T+C/G](image)

Skew: $\frac{A - T}{A + T} + \frac{C - G}{C + G}$
A wavelet based methodology to detect replication origins

Experimentally observed replication origin in the human genome

Lamin B2: 2368 kb

Chromosome 19

Predicted RO: 2365 kb

Skew: \[ \frac{A - T}{A + T} + \frac{C - G}{C + G} \]
Transcription bias
Transcription bias

Detecting discontinuities using the wavelet transform
Application to a known human replication origin

C-MYC origin (chromosome 8)

First evidence of a replication bias in human DNA
Application to a known human replication origin

C-MYC origin (chromosome 8)

First evidence of a replication bias in human DNA

Chromosome 21
Application to a known human replication origin

C-MYC origin (chromosome 8)

First evidence of a replication bias in human DNA

Our model: well defined replication origins, separated by diffuse terminuses
Profile detection using an analyzing wavelet adapted to the shape of replicons

C-MYC origin (chromosome 8)
Profile detection using an analyzing wavelet adapted to the shape of replicons

C-MYC origin (chromosome 8)

Chromosome 21
Deterministic Chaos in DNA Sequences

- **Human Chromosome**
  - 22
  - 11

**Phase Portrait**
- Genes: anti-sense, sense, inter

**Poincare Map**

**1D Map**

**Shil’nikov chaotic oscillator**
\[
\dddot{x} + \ddot{x} + \mu_1 \dot{x} + \mu_0 x = -x^3
\]
\[
\mu_0 = -5.5, \mu_1 = 3.5
\]

**Uncorrelated random walk**
SHIL’NIKOV HOMOClinIC CHAOS

(a) Phase portrait

(b) Homoclinic orbit

(c) Poincaré map

(d) 1D map
LYAPUNOV EXPONENTS

$S_{AT} - S_{GC}$ skew profiles smoothed at scale 160 kb

<table>
<thead>
<tr>
<th></th>
<th>d</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome 11 (24Mb)</td>
<td></td>
<td>12.6</td>
<td>8.9</td>
<td>6.1</td>
<td>6.9</td>
<td>7.9</td>
</tr>
<tr>
<td>(NT_033899.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromosome 14 (68Mb)</td>
<td></td>
<td>15.0</td>
<td>10.2</td>
<td>8.8</td>
<td>8.7</td>
<td>10.4</td>
</tr>
<tr>
<td>(NT_026437.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromosome 21 (29Mb)</td>
<td></td>
<td>12.2</td>
<td>8.7</td>
<td>7.4</td>
<td>8.6</td>
<td>11.3</td>
</tr>
<tr>
<td>(NT_011512.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromosome 22 (23Mb)</td>
<td></td>
<td>12.5</td>
<td>8.1</td>
<td>6.3</td>
<td>5.8</td>
<td>7.2</td>
</tr>
<tr>
<td>(NT_011520.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shil’nikov strange attractor (30Mb)</td>
<td></td>
<td>4.2</td>
<td>5.6</td>
<td>6.5</td>
<td>7.3</td>
<td>7.1</td>
</tr>
</tbody>
</table>

Computation of the largest Lyapunov exponent ($\times 10^3$) using the TISEAN package for a time delay $\tau = 60$ kb and an embedding dimension $d$.

Equation of non-linear oscillator which displays homoclinic chaos of Shil’nikov’s type:

$$\ddot{\theta} + \mu_2 \dot{\theta} + \mu_1 \dot{\theta} + \mu_0 \theta + k\theta^3 = 0$$

$\theta$ and $t$ were rescaled so that the chaotic trajectory displays similar amplitude and characteristic frequencies as the skew oscillatory profiles.
Strand Compositional Asymmetry

Chromosome 21 (Human)

\[ T_{g_0}(n, a_2^*) \]

\[ T_{g_0}(n, a_2^*) \]

-heterochromatin
-GC content

\[ \frac{A-T}{A+T} + \frac{C-G}{C+G} \]

-sense genes
-anti-sense genes
-non-coding sequences

Filtering scales: \( a_1^* = 40\text{kb}, a_2^* = 160\text{kb} \)
Phase Portrait Representation of AT+CG skew

Chromosome 21

Chromosome 22

Filtering scale: \( a_2^* = 160 \text{kb} \)
REFERENCES

Transcription-coupled and splicing-coupled strand asymmetries in eukaryotic genomes.

Low Frequency rhythms in human DNA sequences : a key to the organization of gene location and orientation?

From scale invariance to deterministic chaos in DNA sequences : towards a deterministic description of gene organization in the human genome

Transcription-coupled TA and GC strand asymmetries in the human genome.

Long-range correlations between DNA bending sites : relation to the structure and dynamics of nucleosomes.

Long-range correlations in genomic DNA : a signature of the nucleosomal structure.

Nucleotide composition effects on the long-range correlations in human genes.

Wavelet based fractal analysis of DNA sequences.

Characterizing long-range correlations in DNA sequences from wavelet analysis.