

Research article

Open Access

Phylogenetic distribution of large-scale genome patchiness

José L Oliver*¹, Pedro Bernaola-Galván², Michael Hackenberg^{1,3} and Pedro Carpena²

Address: ¹Dpto de Genética, Facultad de Ciencias, Universidad de Granada, Spain, ²Dpto de Física Aplicada II, Universidad de Málaga, Spain and ³Bioinformatics Group, CIC bioGUNE, CIBER-HEPAD, Technology Park of Bizkaia, 48160 Derio, Bizkaia, Spain

Email: José L Oliver* - oliver@ugr.es; Pedro Bernaola-Galván - rick@uma.es; Michael Hackenberg - mlhack@gmail.com; Pedro Carpena - pcarpena@ctima.uma.es

* Corresponding author

Published: 11 April 2008

Received: 15 November 2007

BMC Evolutionary Biology 2008, 8:107 doi:10.1186/1471-2148-8-107

Accepted: 11 April 2008

This article is available from: <http://www.biomedcentral.com/1471-2148/8/107>

© 2008 Oliver et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The phylogenetic distribution of large-scale genome structure (i.e. mosaic compositional patchiness) has been explored mainly by analytical ultracentrifugation of bulk DNA. However, with the availability of large, good-quality chromosome sequences, and the recently developed computational methods to directly analyze patchiness on the genome sequence, an evolutionary comparative analysis can be carried out at the sequence level.

Results: The local variations in the scaling exponent of the Detrended Fluctuation Analysis are used here to analyze large-scale genome structure and directly uncover the characteristic scales present in genome sequences. Furthermore, through shuffling experiments of selected genome regions, computationally-identified, isochore-like regions were identified as the biological source for the uncovered large-scale genome structure. The phylogenetic distribution of short- and large-scale patchiness was determined in the best-sequenced genome assemblies from eleven eukaryotic genomes: mammals (*Homo sapiens*, *Pan troglodytes*, *Mus musculus*, *Rattus norvegicus*, and *Canis familiaris*), birds (*Gallus gallus*), fishes (*Danio rerio*), invertebrates (*Drosophila melanogaster* and *Caenorhabditis elegans*), plants (*Arabidopsis thaliana*) and yeasts (*Saccharomyces cerevisiae*). We found large-scale patchiness of genome structure, associated with *in silico* determined, isochore-like regions, throughout this wide phylogenetic range.

Conclusion: Large-scale genome structure is detected by directly analyzing DNA sequences in a wide range of eukaryotic chromosome sequences, from human to yeast. In all these genomes, large-scale patchiness can be associated with the isochore-like regions, as directly detected *in silico* at the sequence level.

Background

As soon as genome sequences of sufficient length were available, three groups [1-3] independently described powerful methods (power spectra, analysis of fluctuations in DNA walks) to study large-scale genome structure at sequence level. The emerging view was the existence of

long-range, power-law correlations, thus pointing to fractal (scale-invariant) structure in DNA sequences. However, such fractal structure, implying the existence of DNA segments of all sizes, directly clashes with the view of the genome as composed of long, homogeneous segments (isochores).

Isochores - long ($\gg 300$ kb), compositionally fairly homogeneous genome regions of different average GC levels were uncovered by analytical ultracentrifugation of bulk DNA [4-10]. The phylogenetic distribution of isochores was traditionally studied by centrifugation techniques [7-11], but the analysis of base composition at third codon position or the comparison of GC content between coding and non-coding sequences [12-17] has been also used.

The paradox between a fractal (scale-invariant) or an isochore structure for the genome has been recently solved in the human genome by the discovery that correlations can show deviations from the power-law behavior [18]. Interestingly, such deviations can be associated to isochore-like regions [19] -long-homogeneous genome regions computationally predicted by directly examining the genome sequence and sharing many compositional and biological features with true isochores [20-24].

In this way, the phylogenetic distribution of large-scale genome patchiness can be now explored by analyzing the deviations of power-law behavior in long-range correlations. The method of choice is Detrended Fluctuation Analysis or DFA; the deviations from the power-law can then be revealed by the variations in the local behavior of the scaling exponent α [18,19]. Here, we determined the variation of α at different scales in a wide phylogenetic range of genome sequences. Our analysis clearly distinguishes two characteristic length scales, the larger of which is demonstrated to be unambiguously associated with the isochore-like regions, as detected *in silico*. The phylogenetic distribution of such patterns leads to insights in understanding the evolution of genome compositional heterogeneity.

Results

Two characteristic length scales in human DNA

Long-range correlations are detected by log-log plots of the fluctuation function $F(\ell)$ vs. the length scale ℓ . The scaling exponent α is then given by the slope of the linear fit (see Methods). Figures 1 and 2 show how $\alpha(\ell)$ profiles are used to uncover characteristic scales in human DNA sequences. Significant deviations from the power-law behavior reveal two main characteristics scales in a large human sequence contig (chromosome IV, positions 75671304-167795054, 92.1 Mb). The intermediate scale goes from $\log_{10} \ell > 1.5$ ($\ell > 30$ bp) to $\log_{10} \ell \approx 4.5$ ($\ell \approx 30$ kb), while the large scale spans up to $\log_{10} \ell \approx 7$ ($\ell \approx 10$ Mb). These values are representative for the entire human complement, but when the available contigs were shorter, the large scale typically extended only until $\log_{10} \ell \approx 6$ ($\ell \approx 1$ Mb).

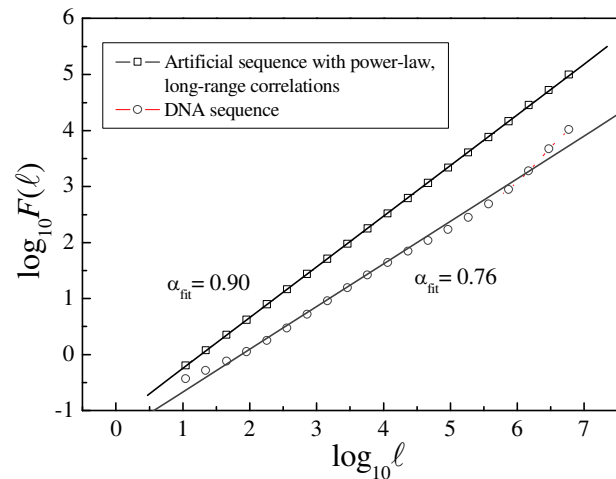


Figure 1

Detecting long-range correlations through DFA. Log-log plots of the fluctuation function $F(\ell)$ vs. the length scale ℓ . The scaling exponent α is given by the slope of the linear fit. The artificial sequence was generated using a standard method to create long-range correlated sequences [31] by imposing that $\alpha = 0.90$, and the DFA recovers correctly this scaling exponent. The DNA sequence analyzed is the largest contig from human chr. IV mapped into a binary sequence using the SW (strong-weak) mapping rule: C or G \rightarrow 1, A or T \rightarrow 0. Its scaling exponent is $\alpha_{\text{fit}} = 0.76$.

Biological source for the two characteristic scales in human DNA

Short- and long-range genome structure is shaped by a wide range of structural, functional, and evolutionary factors acting on the DNA sequence: mutational bias, transcriptional activity, translational constraints, patterns of gene expression, open or closed state of the chromatin, insertion of repeat elements, etc. (see [7-10] for review). Because of its overlap, discerning the role of each of these factors is a difficult task. Experiments of selective shuffling of some genome regions, while leaving others intact, can be used to approach this problem. Here, we used such selective shuffling to identify the biological source for the two characteristic scales identified in DNA sequences [19]. As an example, Figure 3 shows the results for human chromosome 21. When the sequence segments corresponding to the isochore-like regions (as determined by the algorithm IsoFinder [24]) were internally shuffled, the small scale properties disappeared, and the $\alpha(\ell)$ profile remained flat with a constant value of 0.5, as expected in a random sequence without compositional structure. The only effect of the shuffling in the large scale was a slight increase in the corresponding peak: as the shuffling homogenizes the small scale, the large scale patchiness becomes more discernible. Therefore, the large scale properties observed in the $\alpha(\ell)$ profile can be unambiguously

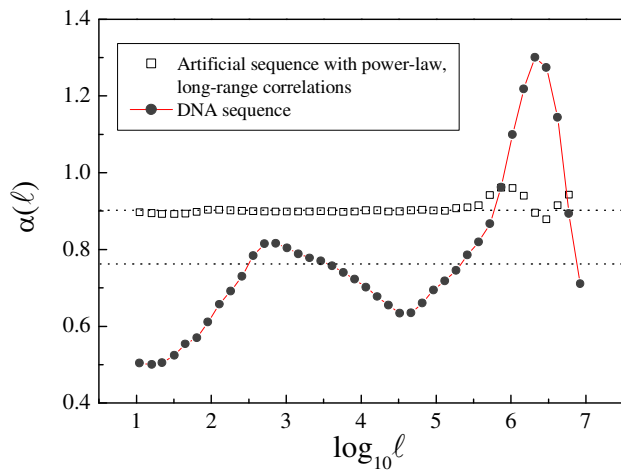


Figure 2
Detecting deviations from power-law behavior. The deviations from power-law behavior can be detected by plotting the local α value vs. the scale ℓ (see Methods). The sequence with real scaling presents a constant value of $\alpha(\ell)$ with small fluctuations around the linear fitting value $-\alpha_{\text{fit}} = 0.90$ (dotted line). However, in human DNA, $\alpha(\ell)$ is far from being a constant: the fluctuations around the value resulting from the linear fitting ($\alpha_{\text{fit}} = 0.76$, dotted line) are not only very large but also following a particular pattern, indicating that this α_{fit} value is meaningless, and that there is no real scaling despite the good linear fitting.

attributed to the computationally predicted, isochore-like regions [19]. However, when the isochore-like sequences were shuffled internally but transposable elements (TEs) were left untouched, or when only the TEs were shuffled internally, the $\alpha(\ell)$ profile is almost identical to the original $\alpha(\ell)$ profile of the natural sequence. This means that the TEs are the main contributors to the heterogeneity at intermediate scale, although introns and exons also make a contribution (see [19] for details).

Phylogenetic distribution

As shown above, the $\alpha(\ell)$ profile revealed as a simple and powerful method to determine the phylogenetic distribution of large-scale patchiness. Two $\alpha(\ell)$ profiles have been already shown for the human genome in Figures 2 and 3. We have also plotted sample $\alpha(\ell)$ profiles for good-quality chromosome sequences from chimpanzee and dog (Fig. 4), mouse and rat (Fig. 5), chicken and *Danio* (Fig. 6), *Drosophila* and *Caenorhabditis elegans* (Fig. 7) and *Arabidopsis thaliana* and *Saccharomyces cerevisiae* (Fig. 8). The numerical values to trace the $\alpha(\ell)$ profiles for the remaining chromosomes with sequence contigs of good-quality and sufficient length from all these species are given in the Additional File 1.

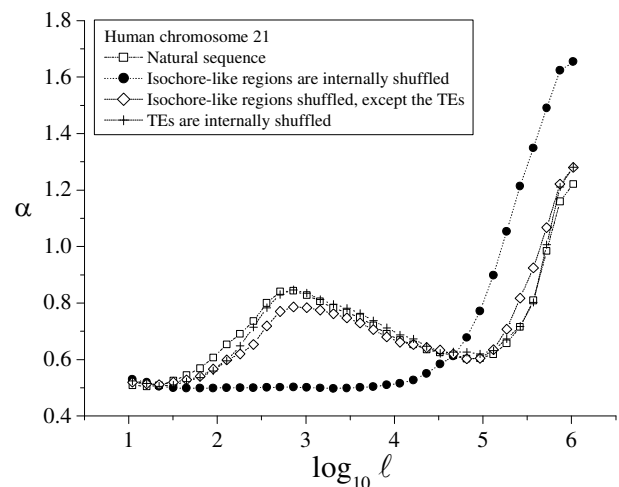


Figure 3
 $\alpha(\ell)$ profile of the human chromosome 21. Plot of the local α value vs. the scale ℓ – the $\alpha(\ell)$ profile – in the human chromosome 21. The profiles for the natural sequence, the artificial sequence obtained after the *in silico* determined isochore-like regions were internally shuffled, the artificial sequence obtained by shuffling the isochore-like regions without touching the TEs, and the artificial sequence obtained by shuffling the TEs are shown.

The extension of the intermediate scale in the other vertebrate genomes is similar to that observed in the human genome, and also in the *Arabidopsis* genome, while in the invertebrates (*Drosophila* and *Caenorhabditis*) the intermediate scale reaches only to 3 kbp. Notably, in the yeast genome this intermediate scale does not exist at all. The beginning of the large scale in all the genomes analyzed ranged from 10 to 30 kbp, while the ending length often extended beyond 100 kbp, or even 1 Mb, depending mainly on the length of the available sequence contigs.

The phylogenetic distribution of compositional patchiness within vertebrates has been traditionally assessed by ultracentrifugation of bulk DNA [11,13,16], uncovering a clear isochore structure in mammals and birds [6,10] and also in some reptiles [16]. The $\alpha(\ell)$ profiles used here reveal large-scale genome patchiness on a wider phylogenetic range, extending to invertebrates (*Drosophila* and *C. elegans*), plants (*A. thaliana*) and yeasts (*S. cerevisiae*).

Besides the profiles for the natural sequence, all these figures also include the profiles for the corresponding artificial sequence obtained by internally shuffling the isochore-like sequence regions, as predicted by IsoFinder [24]. In all the genomes analyzed, the large scale was practically unaffected (save the peak increase due to the homogenization at the small scales). Therefore, the large-

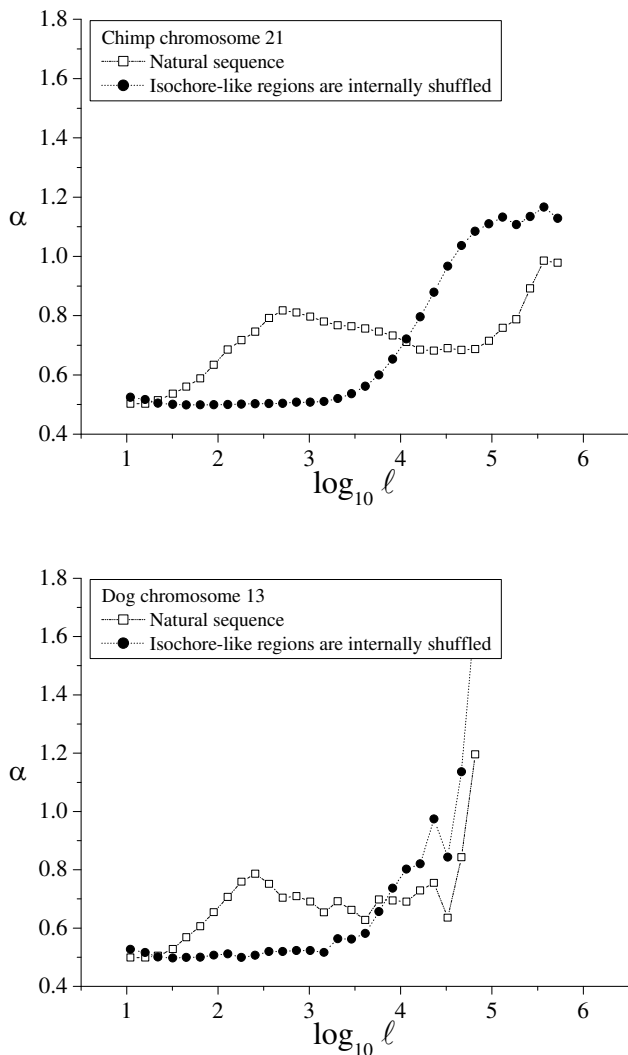


Figure 4
 $\alpha(\ell)$ profiles in chimpanzee and dog. The $\alpha(\ell)$ profiles in chimpanzee and dog genomes. The profiles for the natural sequence of an example chromosome from each genome, and the corresponding artificial sequence obtained after the isochore-like regions were internally shuffled, are shown.

scale patchiness observed in the $\alpha(\ell)$ profiles in all these species can be attributed to the *in silico* determined, isochore-like regions.

Discussion

Large-scale genome compositional heterogeneity has been traditionally revealed through analytical ultracentrifugation of bulk DNA in a large range of genomes (for recent reviews see [8,10]). However, at the sequence level, such demonstration has been hampered by the lack of a consistent, reliable, and widely applicable method. The $\alpha(\ell)$ profiles used here prove to be an excellent tool for

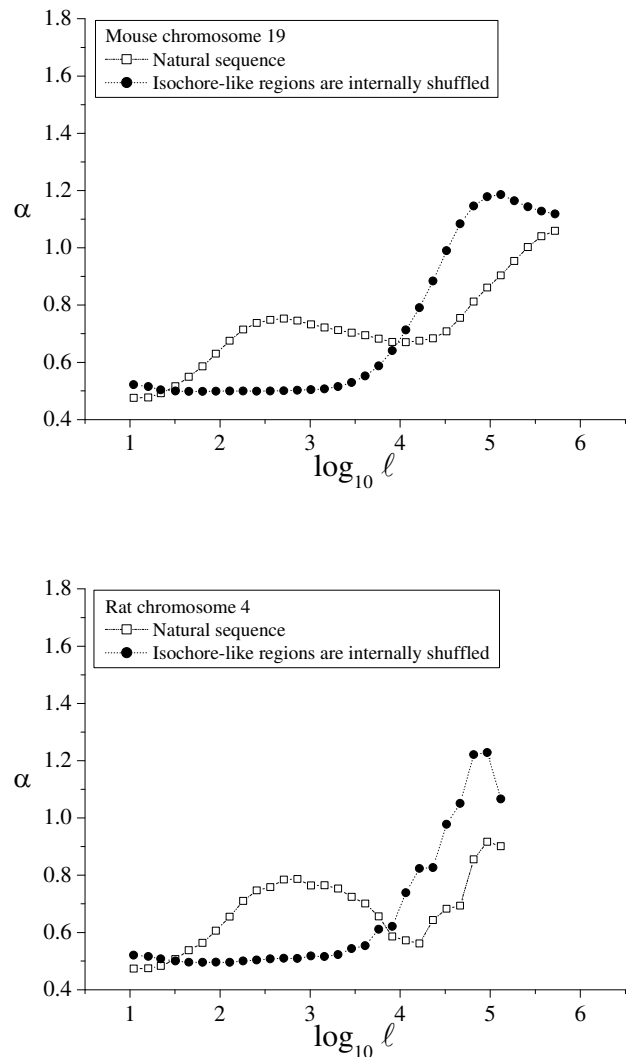


Figure 5
 $\alpha(\ell)$ profiles in mouse and rat. The $\alpha(\ell)$ profiles in mouse and rat genomes. See the legend of Figure 4 for additional details.

this task, allowing us to determine the extent of the different scales of genome compositional heterogeneity in a wide phylogenetic range: yeasts, plants, invertebrates, fishes, birds and mammals. Additionally, the selective shuffling experiments used here to specifically randomize selected genome segments, while not touching others, allow us to identify the *in silico* determined, isochore-like regions as the biological source for large-scale patchiness throughout the entire range of the species analyzed.

A clear limitation of our approach is that it depends critically on the availability of good-quality sequence contigs of sufficient length. For this reason, we have analyzed only the best genome assemblies from eleven eukaryotic spe-

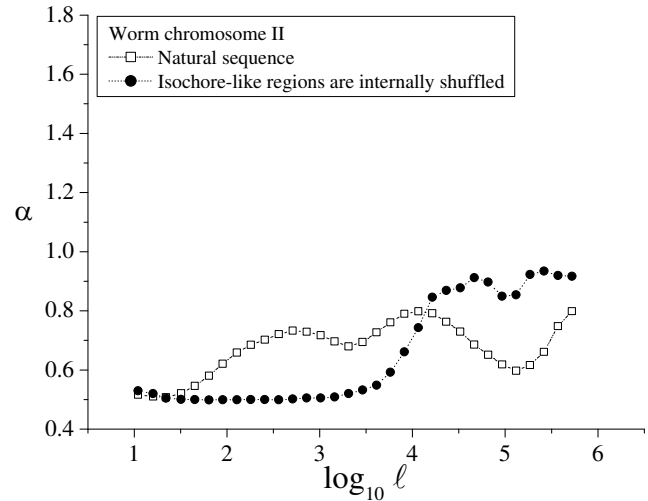
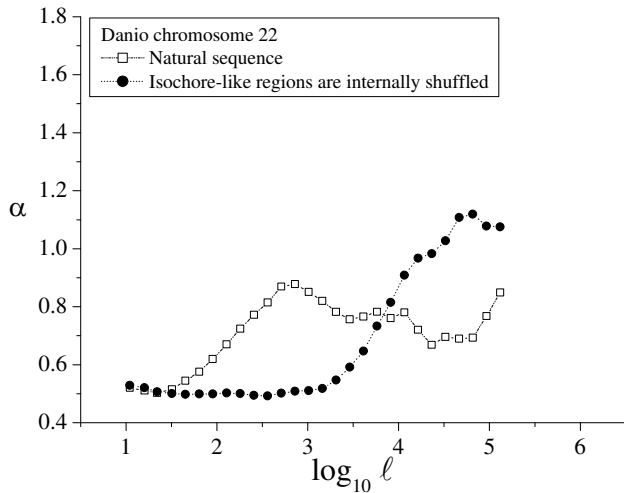
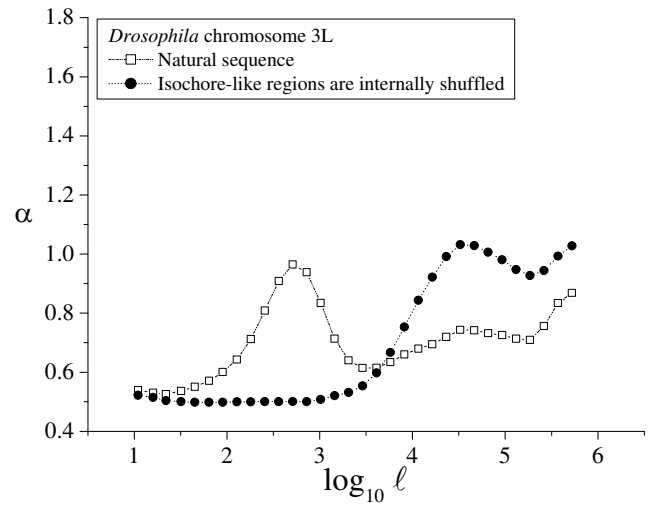
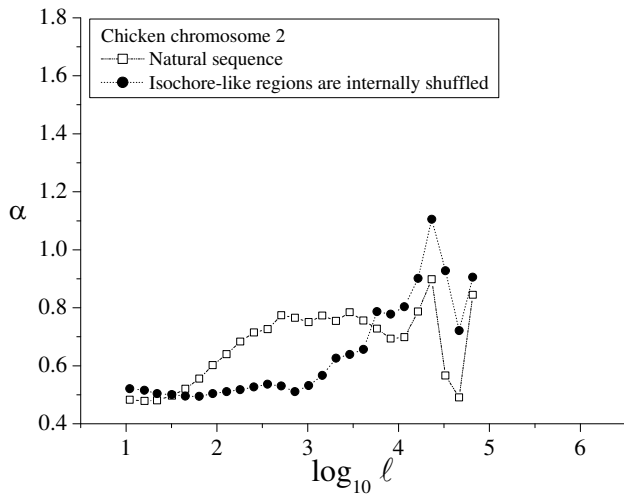


Figure 6
 $\alpha(\ell)$ profiles in birds and fishes. The $\alpha(\ell)$ profiles in chicken and *Danio* genomes. See the legend of Figure 4 for additional details.

Figure 7
 $\alpha(\ell)$ profiles in invertebrates. The $\alpha(\ell)$ profiles in *Drosophila* and worm genomes. See the legend of Figure 4 for additional details.

cies. However, as these species cover a wide phylogenetic spectrum (mammals, birds, fishes, invertebrates, plants and yeasts) the conclusions should be general.

In all these genomes, the large-scale patchiness revealed by $\alpha(\ell)$ profiles can be associated with the isochore-like genome regions predicted by IsoFinder, thus emphasizing the reliability of this algorithm in predicting isochore-like regions at the sequence level.

Conclusion

The analysis of the deviations in the power-law behavior of long-range correlations, through $\alpha(\ell)$ profiles, allowed us to uncover large-scale genome structure in the eleven

sequenced genomes for which sequence contigs of sufficient length and quality are available. Furthermore, through selective shuffling experiments, we were able to identify the computationally-determined, isochore-like structure of these genomes as the biological source for such large-scale patchiness.

Methods

Genomes analyzed

The following genome assemblies, all having sequence contigs of sufficient length, were downloaded from the UCSC Genome Bioinformatics site [25]: *Homo sapiens* (hg18), *Pan troglodytes* (panTro2), *Mus musculus* (mm8),

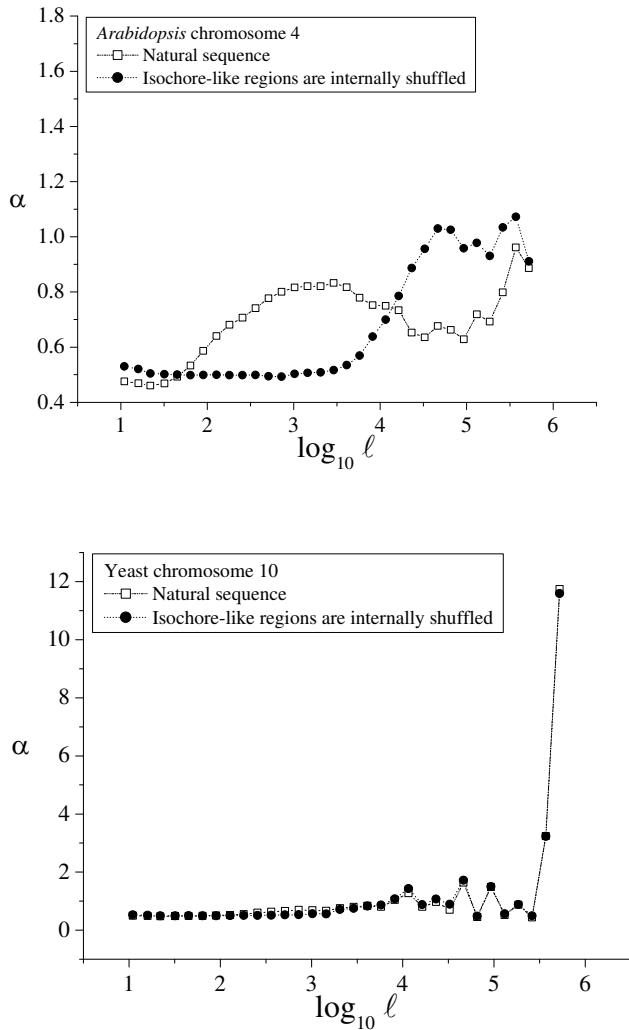


Figure 8
 $\alpha(\ell)$ profiles in plants and yeasts. The $\alpha(\ell)$ profiles in *Arabidopsis* and yeast genomes. See the legend of Figure 4 for additional details.

Rattus norvegicus (rn4), *Canis familiaris* (canFam2), *Gallus gallus* (galGal3), *Danio rerio* (danRer4), *Drosophila melanogaster* (dm2), *Caenorhabditis elegans* (ce2), and *Saccharomyces cerevisiae* (sacCer1). The genome assembly for *Arabidopsis thaliana* (Arab) was downloaded from the NCBI FTP site [26].

Detrended Fluctuation Analysis (DFA) and deviations from perfect power-law behavior

The DFA method [2,27,28] is aimed to detect and quantify long-range correlations in numerical sequences, and therefore DNA sequences have to be mapped into numerical ones prior to the application of DFA. Thus, a DNA chain of length N is of the form $s_1 s_2 \dots s_N$, and the s_i values are obtained according to the strong-weak (SW) mapping

rule: C or G $\rightarrow s_i = 1$, A or T $\rightarrow s_i = 0$. Note that this mapping rule is particularly appropriate to analyze genome-wide correlations, since it corresponds to the most fundamental partitioning of the four bases into their natural pairs in the double helix (A-T and G-C). As expected [29], when the purine-pyrimidine mapping rule was assayed, only structures at intermediate, but not at large, scale were detected (not shown).

To apply DFA, first we create the integrated series by accumulating the original one:

$$\gamma(j) = \sum_{i=1}^j [s_i - \bar{s}] \tag{1}$$

where \bar{s} is the global mean:

$$\bar{s} = \frac{1}{N} \sum_{i=1}^N s_i. \tag{2}$$

The integrated series $\gamma(1) \gamma(2) \dots \gamma(N)$ is divided into N_W windows of equal length ℓ , and in each window we fit the integrated series by using a linear fit γ_{fit} (the local trend). Then, we detrend the integrated series by subtracting the local trend obtaining the detrended fluctuation function Y :

$$Y(i) = \gamma(i) - \gamma_{fit}(i) \tag{3}$$

Finally, we obtain the fluctuation function $F(\ell)$ in this way:

$$F(\ell) = \sqrt{\frac{1}{N_W} \sum_{j=1}^{N_W} \left(\frac{1}{\ell} \sum_{k=1}^{\ell} Y_j^2(k) \right)} \tag{4}$$

where $Y_j(k)$ is the k -th point ($k = 1 \dots \ell$) within the j -th window ($j = 1 \dots N_W$). Therefore, $F(\ell)$ accounts for the average fluctuation of the series around its local trend at scale ℓ .

Fractal long-range correlations appear when $F(\ell)$ scales in the form:

$$F(\ell) \propto \ell^\alpha \tag{5}$$

The exponent α quantifies the type and the strength of the correlations and can be determined by fitting $F(\ell)$ vs. ℓ to a straight line in a log-log plot, the slope being α (Fig. 1). If $\alpha = 0.5$, there is no correlation and the sequence behaves as a random series (white noise), $\alpha < 0.5$ indicates anti-correlations, and $\alpha > 0.5$ positive correlations. However, this procedure can mask information present in the

signal (see [19] for details) and we prefer to determine α as

$$\alpha(\ell) = d \log (F(\ell)) / d \log (\ell)$$

In this way, deviations from power-law behavior lead to local variations of α which can be detected by plotting $\alpha(\ell)$ vs. ℓ (Fig. 2). The DNA sequence shown in this figure, corresponding to the long contig of 92.1 Mb from human chromosome IV, clearly shows two main deviations from the power-law behavior, suggesting the presence of two main characteristic scales (the two major peaks in $\alpha(\ell)$) at intermediate and large ℓ values [19].

Selective shuffling experiments

A shuffled sequence is a random permutation of the original DNA sequence and can be used to test hypotheses concerning any pattern uncovered in DNA sequences. Shuffling can affect the entire sequence or be restricted to only certain regions within a sequence in order to test the influence of such regions on the overall behavior of the one being observed. Here, we carried out selective-shuffling experiments by shuffling only those regions in the genome sequence corresponding to the isochore-like regions predicted by IsoFinder [24]. All the patterns within the isochore-like regions are destroyed by such shuffling, while those out of these regions remain intact. Selectivity in the shuffling process can go a step further when certain genome elements within the isochore-like regions (e.g. TEs) are preserved from shuffling, thereby also remaining intact while the rest of the region is randomized. We obtained both types of partially shuffled sequences in order to identify the biological source for the two length scales observed in the $\alpha(\ell)$ profiles. The coordinates for isochore-like regions identified by IsoFinder on each chromosome sequence of the eleven genomes analyzed in this study are available at the Online Resource on Isochore Mapping [30].

Authors' contributions

PC proposed the idea and, jointly with PBG, wrote the code to locate sequence contigs of sufficient length and compute the $\alpha(\ell)$ profiles. MH wrote the code for selective-shuffling experiments. JLO designed the experimental setup, carried out the computations, wrote the original draft and edited coauthor's contributions. All the authors read and made contributions to the manuscript and approved the final version.

Additional material

Additional file 1

Numerical values for the $\alpha(\ell)$ profiles. The numerical values to trace the $\alpha(\ell)$ profiles for all the chromosomes with sequence contigs of sufficient lengths from the eleven species analyzed here are given (Excel file).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-107-S1.xls>]

Acknowledgements

This work was supported by the Spanish Government (BIO2005-09116-C03-01) and Plan Andaluz de Investigación (CVI-162, P06-FQM-01858, P07-FQM-03163 and TIC-640). The help of David Nesbitt with the English version of the manuscript is also appreciated.

References

- Li W, Kaneko K: **Long-range Correlations and Partial 1/f Spectrum in a Noncoding DNA Sequence.** *Europhysics Letters* 1992, **17**:555-660.
- Peng CK, Buldyrev SV, Goldberger AL, Havlin S, Sciortino F, Simons M, Stanley HE: **Long-range correlations in nucleotide sequences.** *Nature* 1992, **356(6365)**:168-170.
- Voss RF: **Evolution of long-range fractal correlations and 1/f noise in DNA base sequences.** *Phys Rev Lett* 1992, **68(25)**:3805-3808.
- Macaya G, Thierry JP, Bernardi G: **An approach to the organization of eukaryotic genomes at a macromolecular level.** *Journal of molecular biology* 1976, **108(1)**:237-254.
- Thierry JP, Macaya G, Bernardi G: **An analysis of eukaryotic genomes by density gradient centrifugation.** *Journal of molecular biology* 1976, **108(1)**:219-235.
- Bernardi G, Olofsson B, Filipinski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F: **The mosaic genome of warm-blooded vertebrates.** *Science* 1985, **228(4702)**:953-958.
- Bernardi G: **The human genome: organization and evolutionary history.** *Annual review of genetics* 1995, **29**:445-476.
- Bernardi G: **The compositional evolution of vertebrate genomes.** *Gene* 2000, **259(1-2)**:31-43.
- Bernardi G: **Isochores and the evolutionary genomics of vertebrates.** *Gene* 2000, **241(1)**:3-17.
- Bernardi G: **Structural and evolutionary genomics. Natural selection in genome evolution.** Amsterdam, Elsevier; 2004.
- Bucciarelli G, Bernardi G, Bernardi G: **An ultracentrifugation analysis of two hundred fish genomes.** *Gene* 2002, **295(2)**:153-162.
- Barakat A, Matassi G, Bernardi G: **Distribution of genes in the genome of Arabidopsis thaliana and its implications for the genome organization of plants.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95(17)**:10044-10049.
- Bernardi G, Bernardi G: **Compositional patterns in the nuclear genome of cold-blooded vertebrates.** *Journal of molecular evolution* 1990, **31(4)**:265-281.
- Fortes GG, Bouza C, Martinez P, Sanchez L: **Diversity in isochore structure among cold-blooded vertebrates based on GC content of coding and non-coding sequences.** *Genetica* 2007, **129(3)**:281-289.
- Hamada K, Horiike T, Ota H, Mizuno K, Shinozawa T: **Presence of isochore structures in reptile genomes suggested by the relationship between GC contents of intron regions and those of coding regions.** *Genes & genetic systems* 2003, **78(2)**:195-198.
- Hughes S, Clay O, Bernardi G: **Compositional patterns in reptilian genomes.** *Gene* 2002, **295(2)**:323-329.
- Chojnowski JL, Franklin J, Katsu Y, Iguchi T, Guillette LJ Jr., Kimball RT, Braun EL: **Patterns of vertebrate isochore evolution revealed by comparison of expressed Mammalian, avian, and**

- crocodilian genes.** *Journal of molecular evolution* 2007, **65(3)**:259-266.
18. Viswanathan GM, Buldyrev SV, Havlin S, Stanley HE: **Quantification of DNA patchiness using long-range correlation measures.** *Biophysical journal* 1997, **72(2 Pt 1)**:866-875.
 19. Carpena P, Bernaola-Galván, P., Coronado, A.V., Hackenberg, M., Oliver, J.L.: **Identifying characteristic scales in the human genome.** *Phys Rev E* 2007, **75**:32903.
 20. Nekrutenko A, Li WH: **Assessment of compositional heterogeneity within and between eukaryotic genomes.** *Genome Res* 2000, **10(12)**:1986-1995.
 21. Oliver JL, Bernaola-Galvan P, Carpena P, Roman-Roldan R: **Isochore chromosome maps of eukaryotic genomes.** *Gene* 2001, **276(1-2)**:47-56.
 22. Oliver JL, Carpena P, Roman-Roldan R, Mata-Balaguer T, Mejias-Romero A, Hackenberg M, Bernaola-Galvan P: **Isochore chromosome maps of the human genome.** *Gene* 2002, **300(1-2)**:117-127.
 23. Zhang CT, Zhang R: **An isochore map of the human genome based on the Z curve method.** *Gene* 2003, **317(1-2)**:127-135.
 24. Oliver JL, Carpena P, Hackenberg M, Bernaola-Galvan P: **IsoFinder: computational prediction of isochores in genome sequences.** *Nucleic acids research* 2004, **32**:W287-W292.
 25. **The UCSC Genome Bioinformatics** [<http://hgdownload.cse.ucsc.edu/downloads.html>]
 26. **NCBI FTP site** [<ftp://ftp.ncbi.nih.gov/genomes/>]
 27. Coronado AV, Carpena P: **Size effects in correlation measures.** *Journal of Biological Physics* 2005, **31**:121-133.
 28. Peng CK, Buldyrev SV, Goldberger AL, Havlin S, Sciortino F, Simons M, Stanley HE: **Fractal landscape analysis of DNA walks.** *Physica A* 1992, **191(1-4)**:25-29.
 29. Bernaola-Galvan P, Carpena P, Roman-Roldan R, Oliver JL: **Study of statistical correlations in DNA sequences.** *Gene* 2002, **300(1-2)**:105-115.
 30. **Online Resource on Isochore Mapping** [<http://bioinfo2.ugr.es/isochores/>]
 31. Makse HA, Havlin S, Schwartz M, Stanley HE: **Method for generating long-range correlations for large systems.** *Physical Review E Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 1996, **53(5)**:5445-5449.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

