UACM

Universidad Autónoma
de la Ciudad de México

NADA HUMANO ME ES AJENO

COLEGIO DE CIENCIAS Y HUMANIDADES

MAESTRÍA EN CIENCIAS DE LA COMPLEJIDAD

Frequency chaos game method and fractals show evolutionary relationships of the PRKN gene in primates

PUBLICACIÓN ARBITRADA QUE PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS DE LA COMPLEJIDAD

PRESENTA

Rodrigo de Jesús Pérez Gala

Lectores

Mtro. José Luis Gutiérrez Sánchez

Dr. Fernando Ramírez Alatriste

Ciudad de México, octubre 2025.

UACM
Universidad Autónoma de la Ciudad de México
Nada humano me es ajeno

CA Coordinación Académica

SISTEMA BIBLIOTECARIO DE INFORMACIÓN Y DOCUMENTACIÓN

UNIVERSIDAD AUTÓNOMA DE LA CIUDAD DE MÉXICO
COORDINACIÓN ACADÉMICA

RESTRICCIONES DE USO PARA LAS TESIS DIGITALES

DERECHOS RESERVADOS©

CHAOS
Theory and Applications
in Applied Sciences and Engineering

e-ISSN: 2687-4539
RESEARCH ARTICLE
Vol.x / No.x / 2024 / pp.xxx-xxx
https://doi.org/10.51537/chaos.xxxxxxx

Frequency chaos game method and fractals show evolutionary relationships of the *PRKN* gene in primates

Rodrigo Perez-Gala[*,1], **Yordy S. Cangrejo-Useda**[α,2], **Camila V. Castillo-Lopez**[α,3], **Maria J. Sanchez-Manjarrez**[α,4], **Cristian E. Cadena-Caballero**[α,5], **Francisco Martinez-Perez**[αβσ,6] **and Fernando Ramirez-Alatriste**[*,7]

*Maestría en Ciencias de la Complejidad, Universidad Autónoma de la Ciudad de México Campus Del Valle, Ciudad de México, 03100, México., [α] Laboratorio de Genómica Celular Aplicada (LGCA), Grupo de Investigación en Cómputo Avanzado y a Gran Escala (CAGE), Universidad Industrial de Santander, Bucaramanga, Santander, 680006, Colombia. , [β] Grupo de Investigación, en Microbiología y Genética. Universidad Industrial de Santander, Bucaramanga, Santander, 680006, Colombia., [σ] Centro de Supercomputación y Cálculo Científico de la Universidad Industrial de Santander – (SC3UIS) Universidad Industrial de Santander, Bucaramanga, Santander, 680006, Colombia.

ABSTRACT The Chaos Game Representation (CGR) algorithm and its frequency-based optimization, the Frequency Chaos Game Representation (FCGR), offer alignment-free methods for analyzing DNA sequences through fractal geometry. This study investigates the evolutionary relationships of the *PRKN* gene in primates using FCGR, exploring his capacity to reveal phylogenetic signals. We applied FCGR to *PRKN* gene sequences from 16 primate species, calculating nucleotide frequencies and generating fractal representations. Phylogenetic relationships were inferred from fractal similarity and compared to established phylogenies and Shannon entropy was employed to correlate sequence organization with fractal patterns. Results demonstrate that FCGR effectively captures evolutionary relationships of the *PRKN* gene, yielding phylogenetic clustering consistent with conventional methods. The fractal patterns and their relation to Shannon entropy reveal structural organization within the *PRKN* gene sequence, independent of sequence length. This alignment-free, fractal-based approach offers a rapid and informative tool for studying genetic evolution, with potential applications in understanding primate phylogeny and neurodegenerative disorders linked to *PRKN*.

KEYWORDS
Chaos Game Representation Frequency
Chaos Game Representation Fractals
Phylogeny
PRKN gene
Primates
Evolutionary Relationships

INTRODUCTION

Complex systems science provides a powerful framework for understanding biological organization across multiple scales, from molecules to ecosystems Siegenfeld and Bar-Yam (2020). While a universal definition remains elusive, complex systems are characterized by large ensembles of interacting elements, spontaneous self-organization, and emergent non-trivial structures that cannot be predicted solely from individual components Mitchell and Newman (2002). Within this paradigm, fractal geometry, the study of self-similar patterns at different scales, offers a unique lens for examining biological complexity Retnaningsih (2024). Fractals, characterized by self-similarity, fractional dimensions, and iterative generation, are observed both geometrically and statistically in nature Chatterjee and Yilmaz (1992); Kantelhardt (2008), and increasingly recognized as inherent properties of biological sequences and evolutionary processes Saeed (2020).

The Chaos Game Representation (CGR) is a prime example of applying fractal concepts to DNA sequence analysis. CGR transforms nucleotide sequences into fractal images, notably the Sierpiński triangle, by iteratively mapping sequence elements to defined vertices in a geometric space Barnsley (1988). This method, when applied to DNA, generates a 'Sierpiński carpet'-like fractal, visually encoding sequence-specific patterns Jeffrey (1990). To en-

Manuscript received: 3 November 2023,
Revised: 28 December 2023,
Accepted: 20 January 2024.

[1]galagro35@gmail.com
[2]stivencu2411@gmail.com
[3]taini0623@gmail.com
[4]mariasanchezmanjarres@gmail.com
[5]cadena.9605@gmail.com
[6]fjmartin@uis.edu.co
[7]fernando.ramirez@uacm.edu.mx (**Corresponding author**).

hance the analysis of large DNA sequences, the Frequency Chaos Game Representation (FCGR) was developed. FCGR utilizes k-mer frequencies to generate fractals, offering a more generalized and computationally efficient approach for analyzing long genomic regions Bai-lin *et al.* (2000); Deschavanne *et al.* (1999). These fractal-based methods offer a graphical representation of global sequence properties Löchel and Heider (2021) and have proven effective in phylogenetic analysis without sequence alignment, even for megabase-sized genomes Jeffrey (1990); Bai-lin *et al.* (2000); Deschavanne *et al.* (1999). Furthermore, fractal structure can be quantitatively linked to Shannon entropy Allen *et al.* (2009), a measure of information content or disorder within a sequence. By quantifying entropy, we can assess the information distribution and identify recurring patterns reflected in fractal formation Allen *et al.* (2009), providing insights into the intrinsic organization of genetic information as a complex system Mouchet and Mouillot (2010).

The study of complex patterns in biology has been explored across various disciplines, including physiology through chaos theory Boubaker (2024), as well as in genetic networks exhibiting chaotic behavior Kozlovska and Sadyrbaev (2024) and in cryptographic applications where dynamic DNA coding based on chaos enhances image encryption security Patidar and Kaur (2024). Building upon these principles, this study examines the *Parkin RBR E3 Ubiquitin Protein Ligase* (*PRKN*) gene, a particularly relevant candidate for fractal analysis due to its inherent complexity. *PRKN* is the second largest gene in the human genome Tanaka (2020), conserved across diverse species including primates Marín *et al.* (2004), and characterized by a large size (1.3 Mbp in primates) with extensive intronic regions exceeding exon size Munk *et al.* (2021). The encoded protein contains multiple conserved domains, including an N-terminal ubiquitin-like domain and C-terminal RING/IBR/RING finger motifs Wang *et al.* (2023), and is involved in critical cellular processes such as protein ubiquitination and mitophagy Leduc-Gaudet *et al.* (2022); Wang *et al.* (2023). Dysfunction of *PRKN* is strongly implicated in neurodegenerative disorders, particularly early-onset Parkinson's disease and autosomal recessive juvenile Parkinsonism Ahmad *et al.* (2023); Olszewska *et al.* (2022). Given its complex structure, conserved evolutionary history, and functional significance, we hypothesize that the *PRKN* gene exhibits a discernible "fractal organization" reflecting its evolutionary trajectory. This fractal organization, visualized through FCGR and CGR, could provide a novel alignment-free approach to infer phylogenetic relationships among primates based on the *PRKN* gene.

Therefore, this study aims to: (1) calculate nucleotide frequencies and generate fractal representations of the *PRKN* gene in 16 primate species using FCGR and CGR; (2) construct a phylogeny based on fractal similarity; (3) explore the relationship between Shannon entropy and fractal patterns in *PRKN* sequences; and (4) compare the fractal-based phylogeny with established primate phylogenies. We hypothesize that FCGR and CGR will effectively reveal evolutionary relationships of the *PRKN* gene in primates, offering a rapid and alignment-free method for phylogenetic inference and providing insights into the complex organization and evolution of this critical gene.

METHODS

Database Construction and Nucleotide Frequency Calculation

***PRKN* Gene Sequence Retrieval** *PRKN* gene sequences from 16 primate species were retrieved from the 'Primate *PRKN* gene database' Cangrejo-Useda *et al.* (2025). Species included: *Homo sapiens*, *Gorilla gorilla gorilla*, *Pan paniscus*, *Pan troglodytes*, *Pongo abelii*, *Pongo pygmaeus*, *Macaca fascicularis*, *Macaca mulatta*, *Macaca thibetana thibetana*, *Rhinopithecus roxellana*, *Trachypithecus francoisi*, *Nomascus leucogenys*, *Symphalangus syndactylus*, *Microcebus murinus*, *Callithrix jacchus* and *Lemur catta*. The database was constructed with NCBI GenBank records. For each species, the longest available sequence annotated as the complete *PRKN* gene (including introns and exons) was selected. Sequence accession numbers are provided in Cangrejo-Useda *et al.* (2025).

Sequence Cleaning and Nucleotide Frequency Analysis To ensure analysis of only nucleotide sequences, identification tags, spaces, and any non-nucleotide characters were removed from each *PRKN* gene sequence using the Code FASTA SEQUENCE CLEANER program (GITHUB of the Complexity Science Group: Chaos, Fractals, Nature Applications (CSFANA) Rodrigo-Gala (2025). This program verifies and cleans nucleic acid sequences downloaded from databases. The Code Nucleotide Relative Frequency Visualizer program (NRFV) ((CSFANA) Rodrigo-Gala (2025)) was then used to calculate the total and relative frequency of each nucleotide (Adenine, Thymine, Guanine, Cytosine) for each *PRKN* gene sequence. These frequencies were used for subsequent fractal frequencies analysis.

Chaos Game Representation (CGR) and Frequency Chaos Game Representation (FCGR) Fractal Generation

Algorithm Description Sierpiński Carpet fractals were generated using the 'Chaos Game' algorithm, implemented for *PRKN* gene sequences with the DNA Frequency Fractal Generator program (DFFG) ((CSFANA)(Rodrigo-Gala 2025)). The DFFG program is based on the work of Allen *et al.* (2009) and Cabrera-Becerril and Rayón (2025), and implements the frequency version of CGR as proposed by Deschavanne *et al.* (1999), modified from the original Jeffrey method Jeffrey (1990).

The FCGR algorithm operates by assigning each of the four DNA nucleotides (A, T, G, C) to a vertex of a unit square. For a given DNA sequence, the algorithm iteratively plots points within the square. The first point is typically placed at the center of the square. Subsequent points are generated by the following steps:

1. **Read the next nucleotide** in the DNA sequence. 2. **Identify the vertex** corresponding to that nucleotide (e.g., A=top-left, T=top-right, C=bottom-left, G=bottom-right). 3. **Calculate the midpoint** between the current point and the vertex identified in step 2. 4. **Plot a new point** at this midpoint. 5. **Repeat steps 1-4** for the entire DNA sequence.

In FCGR, instead of plotting each point individually, the algorithm calculates the frequency of k-mers (sequences of length k) within the DNA sequence. This frequency matrix is then used to generate a grayscale image representing the fractal. Higher k-mer frequencies are represented by darker pixels, and lower frequencies by lighter pixels, creating the Sierpiński carpet fractal pattern.

k-mer Size and Fractal Resolution To determine the appropriate fractal resolution and pattern complexity, k-mer sizes ranging from 3 to 11 nucleotides were tested. A k-mer size of 7 nucleotides was chosen for the final analysis as it provided optimal resolution of recursive patterns and sufficient detail in the fractal images without excessive image darkening observed at larger k-mer sizes (≤ 8). Larger k-mer sizes resulted in overly dense fractals that obscured pattern visualization, while smaller k-mer sizes lacked the complexity to effectively differentiate sequences. Furthermore, the quadrant position in the fractals proposed by Bai-lin *et al.* (2000) was used to construct fractals showed in Figure 1.

(a) $K = 1 - mer$

G	C
A	T

(b) $K = 2 - mer$

GG	GC	CG	CC
GA	GT	CA	CT
AG	AC	TG	TC
AA	AT	TA	TT

(c) $K = 3 - mer$

GGG	GGC	GCG	GCC	CGG	CGC	CCG	CCC
GGA	GGT	GCA	GCT	CGA	CGT	CCA	CCT
GAG	GAC	GTG	GTC	CAG	CAC	CTG	CTC
GAA	GAT	GTA	GTT	CAA	CAT	CTA	CTT
AGG	AGC	ACG	ACC	TGG	TGC	TCG	TCC
AGA	AGT	ACA	ACT	TGA	TGT	TCA	TCT
AAG	AAC	ATG	ATC	TAG	TAC	TTG	TTC
AAA	AAT	ATA	ATT	TAA	TAT	TTA	TTT

Figure 1 The configuration of string counters for K values ranging from 1 to 3 within squares of identical dimensions.

Quadrant Adjustment for Fractal Symmetry To optimize fractal symmetry and definition, two quadrant arrangements for nucleotide assignment within the unit square were tested (Figure 2). The default arrangement (Figure 2a) placed Adenine-Thymine in the upper quadrants and Guanine-Cytosine in the lower quadrants. An alternative arrangement (Figure 2b) was tested, swapping to Cytosine-Adenine in the upper and Thymine-Guanine in the lower quadrants. The arrangement in Figure 2b, with complementary bases arranged diagonally and chemical groups (purines and pyrimidines) placed vertically, was chosen for improved visual symmetry and pattern clarity in the generated fractals.

Analysis of Fractal Images and Matrices

Occurrence frequency and composition of nucleotide k-mers for PRKN gene fractals in primates. The reading and counting of combinations for nucleotide k-mers of the *PRKN* gene from primates were generated with the **DFFG** program, to obtain the nucleotide occurrence frequency of each possible k-mer combination. The result was visualized in a grouped manner based on the number of occurrences and in descending order with the K-mer **Frequency-Clustering (KFC) and K-mer Frequency-Rank Plot (KFRP) programs** (CSFANA)(Rodrigo-Gala 2025).

The fractals generated by DFFG were verified using the **DNA Chaos Game Algorithm program (DCGA)** (CSFANA)(Rodrigo-Gala 2025). In this way, the fractal images of the *PRKN* gene in primates were verified.

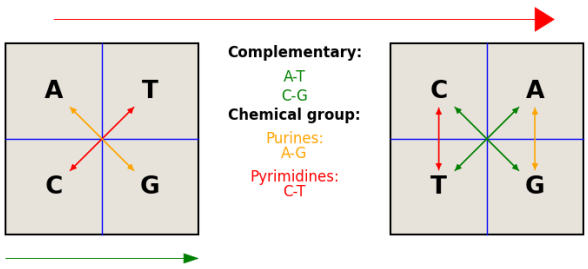


Figure 2 Quadrants arrangements to fractals construction. (a) Default parameters, (b) quadrant adjustment.

Comparative Analysis of Frequency Matrices The DFFG program also generates a frequency matrix representing the k-mer counts for each sequence. To analyze differences between these numerical matrices, the Numerical Matrix Difference Analyzer program (NMDA) ((CSFANA)(Rodrigo-Gala 2025) was employed. NMDA calculates the absolute percentage difference between corresponding cells in two matrices. For a set of matrices $M_1, M_2, ..., M_n$, the difference matrix Δ_i for each matrix M_i compared to all other matrices M_j $(j \neq i)$ is calculated as:

$$\Delta_i = \{|M_i - M_j| \quad | \quad j \neq i\}, \quad \text{for} \quad i = 1, 2, ..., n. \tag{1}$$

The average percentage difference across all comparisons was used as a measure of matrix dissimilarity.

Phylogenetic Analysis and Dendrogram Construction

Phylogenetic Tree Construction For comparison with fractal-based analyses, a phylogenetic tree of the *PRKN* gene was constructed using standard phylogenetic methods. Gene sequences were aligned using the Lamassemble program Frith *et al.* (2021) with default parameters. *Mus musculus PRKN* gene sequence was used as an outgroup to root the tree. Phylogenetic analyses were performed using the CIPRES Science Gateway online platform Miller *et al.* (2015). Maximum likelihood phylogenetic reconstruction was performed using IQ-Tree V.2.1.2 Minh *et al.* (2020) with branch support assessed by ultra-bootstrap Hoang *et al.* (2018) and SH-aLRT Anisimova *et al.* (2011) with 1000 replicates each. The GTR+G substitution model was selected using jModelTest 2 Darriba *et al.* (2012) based on the corrected Akaike information criterion (AICc).

Dendrogram Construction from Fractal Matrices To visualize relationships based on fractal matrix similarity, dendrograms were constructed using the Similarity Dendrogram program (SimDendro) (CSFANA)(Rodrigo-Gala 2025). SimDendro calculates the Euclidean distance between frequency matrices and generates a similarity dendrogram. The cophenetic correlation coefficient was calculated to assess the fit between the dendrogram and the original distance matrix.

Shannon Entropy Calculation

Shannon entropy $(H(X))$ for each *PRKN* gene sequence was calculated using the Shannon Entropy Calculator and Visualizer program (SECV) ((CSFANA)(Rodrigo-Gala 2025)) using the formula:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \cdot \log_2 P(x_i) \quad (2)$$

where $P(x_i)$ is the frequency of each nucleotide (x_i) in the sequence, and $n = 4$ (for A, T, G, C). SECV calculates Shannon entropy and generates a ranked list and graph of entropies for all input sequences.

RESULTS

Primates *PRKN* gene nucleotide characterization

Consistent with previous reports Cangrejo-Useda *et al.* (2025), the *PRKN* gene exhibited a conserved size of approximately 1,300,000 base pairs across the 16 primate species analyzed. Exceptions were observed in *Microcebus murinus* and *Lemur catta*, which displayed smaller sequence sizes, potentially due to intron/exon modifications or deletions (Supplementary Material 1).

The nucleotide composition analysis of the *PRKN* gene revealed an enrichment of adenine-thymine (A-T) compared to cytosine-guanine (C-G) pairs across all species. This pattern aligns with Chargaff's principle of base pair complementarity Vischer and Chargaff (1948) but highlights a specific A-T enrichment in primate *PRKN* sequences. While *M. murinus* and *L. catta* exhibited similar overall patterns, the relative nucleotide frequencies were more balanced in *M. murinus*, potentially reflecting its reduced gene size. In contrast, *L. catta* displayed the highest A-T percentage among all species (supplementary material 1). These structural variations participate in the functional characteristics of the *PRKN* gene Benisty *et al.* (2023) and could be related to the gene expansion and structure Brovkina *et al.* (2023). .

Chaos Game Representation and Sierpiński Carpet Fractals

Nine fractal images, that corresponded to k-mer sizes from 3 to 11, were generated for each species (Supplementary Material 2). Recursive patterns emerged only for k-mer sizes of 7 to 11, with optimal resolution observed at $7-mers$. Larger k-mer sizes ($\geq$ 8) resulted in image darkening, hindering pattern visualization. The fractal images for $7-mers$ displayed asymmetric, grid-like patterns dominated by black and navy-blue squares, reflecting low k-mer presence, particularly for combinations with high C-G content (Supplementary Material 2).

Distinct differences in fractal coloration were observed for *Callithrix jacchus*, *L. catta*, and *M. murinus*. These species exhibited red-dominated quadrants, indicative of altered k-mer distribution. The fractals of *M. murinus* lacked defined patterns, attributed to its reduced sequence size and balanced nucleotide proportions. Meanwhile, the fractals of *L. catta* and *C. jacchus* suggested an influence of tandem repeats within the *PRKN* gene on k-mer distribution Sievers *et al.* (2021).

Adjusting the nucleotide arrangement in fractals (e.g., guanine-adenine and thymine-cytosine quadrants) maintained overall patterns but shifted the location of large squares, further validating the robustness of the methodology (Fig. 2, Supplementary Material 2).

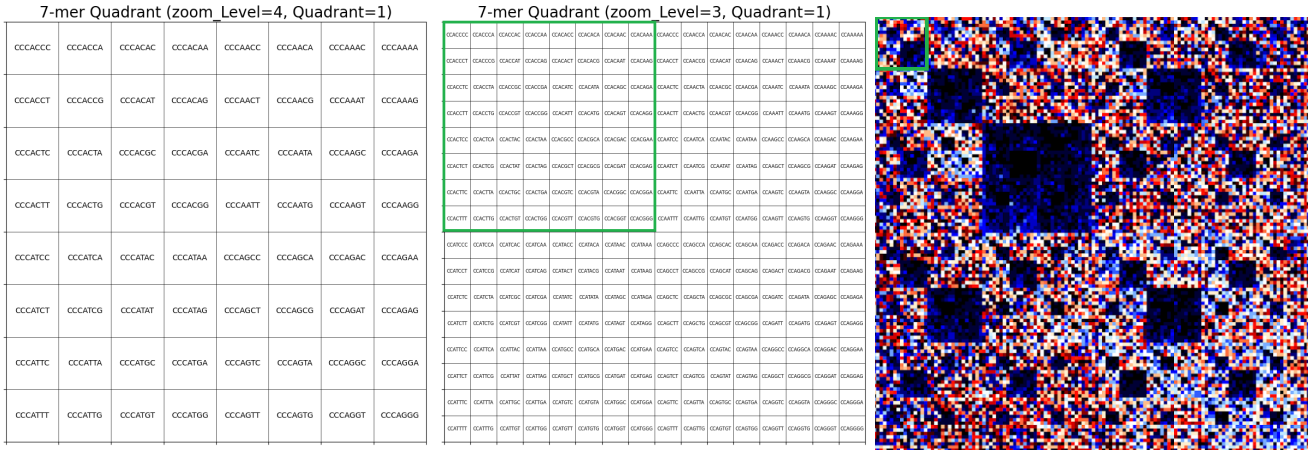
Frequency and Composition of 7-Mer

Analysis of $7-mer$ frequencies revealed a total of 16,384 unique $7-mer$ combinations. Across most species, A/T-rich $7-mers$ occurred more frequently than G/C-rich $7-mers$. However, *Microcebus murinus*, *Callithrix jacchus*, and *Lemur catta* deviated from this trend, exhibiting comparable frequencies of GC-rich and AT-rich $7-mers$ (Quantitative data provided in Supplementary Material 1 and 2). This suggests a shift in k-mer frequency distribution in these species.

This k-mer distribution has showed similarities in the presence of short tandem repeats not only between animals but in the entire eukarya domain, as reported by Sievers *et al.*, (2021), and is higly related with the presence of transposable elements Farré *et al.* (2011); Sievers *et al.* (2021), structures that may be content in the *PRKN* introns and explained: the AT enrichment in the gene and provides a possible explanation of the gene evolution and growth Boissinot (2022); Sievers *et al.* (2021).

Quantitative Comparison of Fractal Images and Matrices

With the CGR and FCGR methodologies, the fractal images for the *PRKN* gene sequences confirmed regions of low and high k-mer density (Supplementary Material 4 and 5). The observed patterns were consistent across species, supporting the validity of



(a) zoom level = 4 **(b)** zoom level= 3 **(c)** without zoom

Figure 3 Visualization of 7-mers at different zoom levels and quadrants. (a) shows a more detailed area with zoom level = 4, where the matrix is divided into 16 parts per side, displaying a specific quadrant. (b) has zoom level = 3, dividing the matrix into 8 parts per side and showing the same quadrant at a larger scale. (c) displays the full matrix without zoom, allowing for structural comparisons at different scales. (illustrates the composition of the fractal.).

the analytical approach. Additionally, dendrograms constructed from Euclidean distance matrices revealed clustering patterns that aligned closely with established phylogenies based on *PRKN* gene sequences and a known primate phylogeny Makova *et al.* (2024); Duda and Zrzavý (2013).

Phylogenetic Relationships from Fractal Analysis and Comparison to Known Phylogeny

Dendrograms constructed from Euclidean distances of frequency matrices (SimDendro) using $7-mer$ data (Figure 4db, 4dc, 4dd) showed phylogenetic clustering largely consistent with established primate phylogenies Makova *et al.* (2024); Duda and Zrzavý (2013) (Supplementary Material 6). Dendrograms without outgroup rooting (Figure 4da) clearly separated Hominoidea and Cercopithecoidea superfamilies. Inclusion of *Mus musculus* as an outgroup (Figure 4db, 4dc) rooted the dendrogram and placed *Microcebus murinus* and *Lemur catta* as distinct outgroups to the main primate cluster, although the exact placement of *Microcebus murinus* varied depending on analysis parameters (Figure 4dc, 4dd). The cophenetic correlation coefficient for the dendrograms was consistently high (≥ 0.90), indicating a good fit between the dendrogram and the original distance matrix. Phylogenetic trees constructed using maximum likelihood (IQ-Tree) (Figure 4db, 4dc, 4dd) largely mirrored the topology of the fractal-based dendrograms, further supporting the phylogenetic signal captured by FCGR.
These results are consistent with previously reported *PRKN* Primate gene phylogeny (Cangrejo-Useda *et al.* 2025), which showed similar cluster of species. This suggests that fractal-based analyses capture meaningful evolutionary signals without the need for alignment. It is noteworthy that those phylogenetical reconstruction did not represented the evolutionary history of Primate species, but generates similar patterns, which suggest a strong conservation in this gene.
However, the inclusion of *PRKN* sequence of *M. musculus* in the dendogram construction, showed a different cluster in primates species, and suggested *M. murinus* as outgroup. This could be due to differences in gene size, which indicates the high influence of the sequence size in the dendogram construction. Despite this, the clustering of species without *M. musculus*, showed similarities respect to the *PRKN* phylogeny constructed with the standard protocol but in a reduced cost in time and computer requeriments, that reflects de accuracy of the method proposed here.

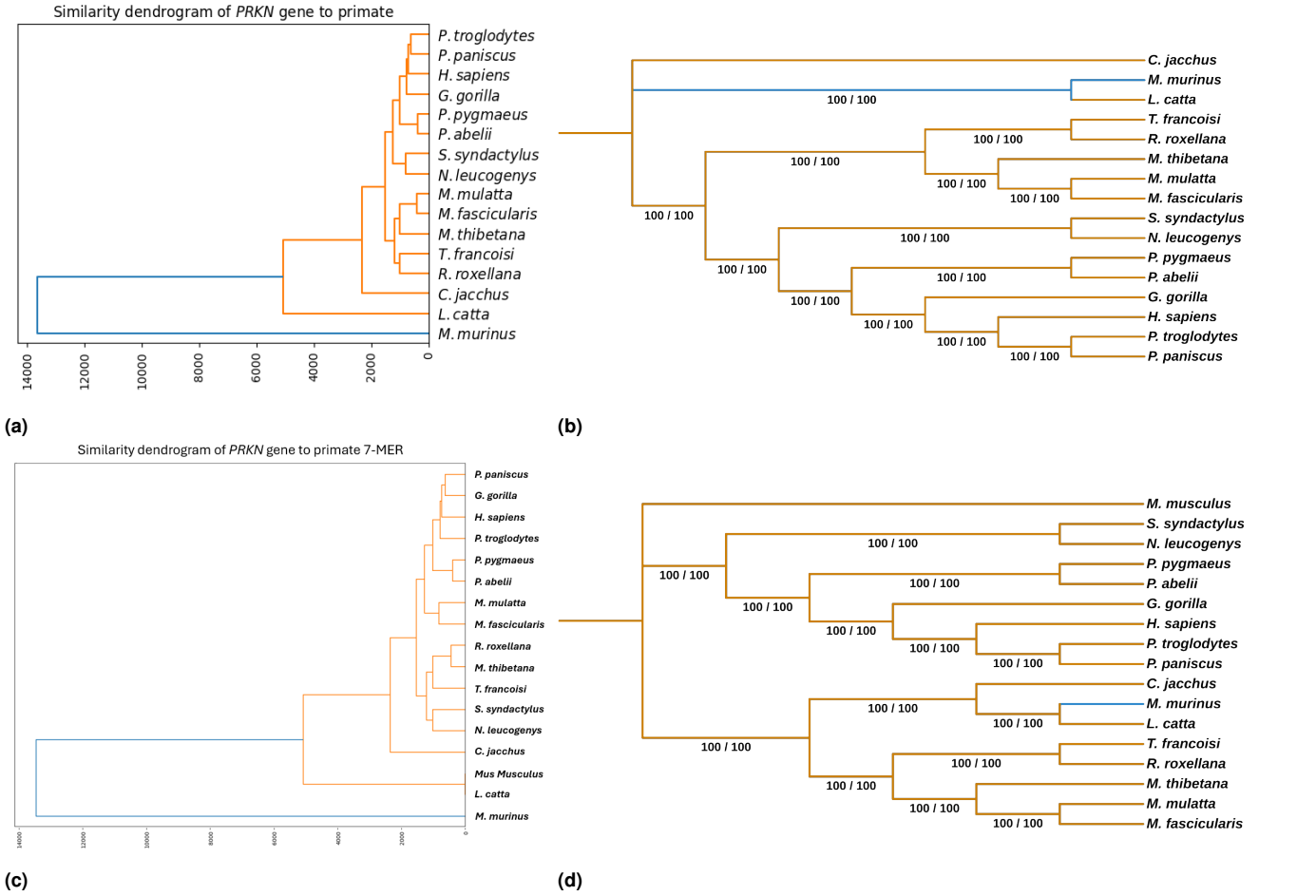


Figure 4 *PRKN* primates gene dendograms and phylogenies. Dendograms were generated with 7-mers and phylogenies were constructed with maximun likelihood aproximation. (a) and (b) corresponded to analisys without outgroup; and (c)-(d) included *M. musculus* as outgroup and rooted the phylogeny. (b) phylogeny was rooted by default. Numbers in phylogenies represents the bootstraps values for each node. Colors in branches indicate the species clustering similarities in dendograms respect to phylogenies. Exceptions were founded in (c) and (d), with diferent clustering to *M. murinus*.

Shannon Entropy of *PRKN* Gene Sequences

Shannon entropy values for *PRKN* gene sequences ranged from 1.96 (*Lemur catta*) to 1.99 (*Microcebus murinus*), with an average of 1.97 for the other 14 species (Quantitative data provided in Supplementary Material 5). Lower Shannon entropy in *Lemur catta* correlated with more uniform, less complex fractal patterns. Conversely, higher entropy in *Microcebus murinus* corresponded to more diffuse, less defined fractal patterns. Overall, a trend was observed: species with lower Shannon entropy tended to exhibit more structured and visually distinct fractal patterns, while higher entropy correlated with less defined fractal structures, suggesting an inverse relationship between sequence randomness (entropy) and fractal pattern organization.

The high entropy in *M. murinus* could be related with its sequence size, however, it is difficult to determined the fidelity of the sequence, based in the gene physiology, its pattern of expression and the diffucult to access to genetic information of the species Ahmad *et al.* (2023); Tanaka (2020). However, was possible to construct a dendogram similar to the phylogeny stablished to *PRKN* gene.

DISCUSSION

This study demonstrates the effectiveness of Frequency Chaos Game Representation (FCGR) and Chaos Game Representation (CGR) methods for analyzing evolutionary relationships of the *PRKN* gene in primates through fractal geometry. Our results show that FCGR, an alignment-free approach, successfully captures phylogenetic signals within *PRKN* gene sequences, yielding phylogenetic clustering largely congruent with established primate phylogenies derived from traditional alignment-based methods Makova *et al.* (2024); Duda and Zrzavý (2013). This congruence is supported by both visual inspection of fractal patterns, quantitative analysis of fractal image and matrix differences, and dendrogram construction based on fractal similarity.

Advantages and Limitations of FCGR for Phylogenetic Analysis

FCGR offers several advantages for phylogenetic analysis, particularly for long DNA sequences like the *PRKN* gene. Being alignment-free, FCGR circumvents the computational cost and potential biases associated with multiple sequence alignment, a significant bottleneck for large-scale genomic datasets. FCGR provides a holistic, global view of sequence organization, capturing patterns that might be missed by alignment-based methods focused on local sequence similarity. The graphical fractal representation facilitates visual comparison of complex sequence patterns across species. However, FCGR also has limitations. As an alignment-free method, it may lack the fine-grained resolution of alignment-based methods for detecting subtle evolutionary differences at the nucleotide level. The choice of k-mer size is crucial and can influence the resulting fractal patterns and phylogenetic inference. Further research is needed to optimize k-mer size selection and explore the sensitivity of FCGR to different evolutionary scenarios.

Sequence Size Influence and *Microcebus murinus* Placement

The placement of *Microcebus murinus* as an outgroup in fractal-based dendrograms, and its variable positioning depending on analysis parameters, raises interesting questions. While consistent with the use of *Mus musculus* as an outgroup in the phylogenetic tree, the smaller sequence size of *Microcebus murinus PRKN* gene compared to other primates might influence fractal pattern generation and dendrogram placement. Smaller sequences may generate less complex or less defined fractal patterns, potentially affecting distance calculations and clustering. Future studies could investigate methods to normalize fractal analysis for sequence length variations, or explore the use of sliding window FCGR approaches to analyze sub-regions of the *PRKN* gene and assess regional variations in fractal patterns and phylogenetic signal.

PRKN Gene Fractal Organization, Entropy, and Evolution

The observed relationship between Shannon entropy and fractal patterns suggests a link between sequence organization and fractal structure in the *PRKN* gene. Lower Shannon entropy, indicative of less random nucleotide distribution, tended to correlate with more structured and visually distinct fractal patterns. Conversely, higher entropy correlated with less defined fractals. This suggests that regions of lower sequence entropy within the *PRKN* gene may contribute disproportionately to the formation of defined fractal patterns, potentially reflecting functionally important or evolutionarily conserved regions. While this study does not directly link fractal patterns to specific functional domains of the *PRKN* protein, future research could investigate the fractal organization of exons and introns separately and explore correlations between fractal patterns, entropy profiles, and known functional elements within the *PRKN* gene. Furthermore, investigating how selection pressures and evolutionary events shape the fractal organization of genes like *PRKN* could provide novel insights into the evolution of gene structure and function.

Limitations and Future Directions

This study provides a preliminary investigation into the application of FCGR and CGR for primate *PRKN* gene phylogeny. Limitations include the use of a single gene (*PRKN*) and a limited number of primate species. Further studies should expand the analysis to include additional genes and a broader taxonomic sampling of primates and other mammals. Exploring the application of FCGR to other types of genomic data, such as non-coding regions or whole genomes, could further demonstrate the versatility of this approach. Developing more sophisticated quantitative metrics for characterizing fractal patterns beyond entry-by-entry differences could also enhance the analytical power of FCGR. Future research could also explore the biological significance of the observed fractal patterns and their relationship to gene function, regulation, and evolutionary adaptation.

CONCLUSIONS

The Frequency Chaos Game Representation (FCGR) method, and its base algorithm CGR, provide a valuable alignment-free approach for analyzing evolutionary relationships of the *PRKN* gene in primates. FCGR effectively captures phylogenetic signals, yielding dendrograms largely congruent with established phylogenies. The study highlights the potential of fractal geometry and FCGR as a rapid and informative tool for phylogenetic inference, particularly for long DNA sequences, and provides novel insights into the complex organization and evolution of the *PRKN* gene. Further research is warranted to explore the full potential of fractal-based methods in genomics and evolutionary biology, and to investigate the biological significance of fractal patterns in gene sequences.

Acknowledgments

This work was supported by Maestria en Ciencas de la Complejidad from the Universidad Autonoma de la Ciudad de Mexico, Campus Colonia del Valle, Mexico; by the Complexity Science Group: Chaos, Fractals, Nature Applications from the Universidad Autonoma de la Ciudad de Mexico. Moreover, by a research stay

granted by the Universidad Industrial de Santander by resolution 558 of October 20, 2023, and for the Grupo de Investigación en Computo Avanzado y a Gran Escala (CAGE) from Centro de Supercomputacion y Calculo Cientifico of the Universidad Industrial de Santander.

Availability of data and material

Supplementary material can be found in (Perez-Gala *et al.* 2025).

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

LITERATURE CITED

Ahmad, A., D. Nkosi, and M. A. Iqbal, 2023 Park2 microdeletion or duplications have been implicated in different neurological disorders including early onset parkinson disease. Genes **14**: 600.

Allen, B., M. Kon, and Y. Bar-Yam, 2009 A new phylogenetic diversity measure generalizing the shannon index and its application to phyllostomid bats. The American Naturalist **174**: 236–243.

Anisimova, M., M. Gil, J. F. Dufayard, C. Dessimoz, and O. Gascuel, 2011 Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Systematic biology **60**: 685–699.

Bai-lin, H., L. H.C., and Z. Shu-yu, 2000 Fractals related to long dna sequences and complete genomes. Chaos, Solitons Fractals **11**: 825–836.

Barnsley, M. F., 1988 *Fractals Everywhere*. Academic Press, San Diego.

Benisty, H., X. Hernandez-Alias, M. Weber, M. Anglada-Girotto, F. Mantica, *et al.*, 2023 Genes enriched in a/t-ending codons are co-regulated and conserved across mammals. Cell systems **14**: 312–323.e3.

Boissinot, S., 2022 On the base composition of transposable elements. International Journal of Molecular Sciences **23**.

Boubaker, O., 2024 Chaos in physiological control systems: Health or disease? Chaos Theory and Applications **6**: 1–12.

Brovkina, M., M. Chapman, M. Holding, *et al.*, 2023 Emergence and influence of sequence bias in evolutionarily malleable, mammalian tandem arrays. BMC Biology **21**.

Cabrera-Becerril, A. and U. Rayón, 2025 Chaos-game_pygame.ipynb. Accedido: 16 de enero de 2025.

Cangrejo-Useda, Y. S., R. Perez-Gala, L. M. Vera-Cala, C. Barrios-Hernandez, E. P. Ardila-Sandoval, *et al.*, 2025 Prkn gene introns in families from order primates show common open reading frames.

Chatterjee, S. and M. R. Yilmaz, 1992 Chaos, fractals and statistics. Statistical Science **7**: 49–68.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada, 2012 jmodeltest 2: more models, new heuristics and parallel computing. Nature Methods 2012 9:8 **9**: 772–772.

Deschavanne, P. J., A. Giron, J. Vilain, G. Fagot, and B. Fertil, 1999 Genomic signature: characterization and classification of species assessed by chaos game representation of sequences. Molecular Biology and Evolution **16**: 1391–1399.

Duda, P. and J. Zrzavý, 2013 Evolution of life history and behavior in hominidae: Towards phylogenetic reconstruction of the chimpanzee–human last common ancestor. Journal of Human Evolution **65**: 424–446.

Farré, M., M. Bosch, F. López-Giráldez, M. Ponsà, and A. Ruiz-Herrera, 2011 Assessing the role of tandem repeats in shaping the genomic architecture of great apes. PLOS ONE **6**: 1–12.

Frith, M. C., S. Mitsuhashi, and K. Katoh, 2021 lamassemble: Multiple alignment and consensus sequence of long reads. Methods in molecular biology (Clifton, N.J.) **2231**: 135–145.

Hoang, D. T., O. Chernomor, A. V. Haeseler, B. Q. Minh, and L. S. Vinh, 2018 Ufboot2: Improving the ultrafast bootstrap approximation. Molecular biology and evolution **35**: 518–522.

Jeffrey, H. J., 1990 Chaos game representation of gene structure. Nucleic Acids Research **18**: 2163–2170.

Kantelhardt, J. W., 2008 Fractal and multifractal time series. Physics Reports **808**: 1–43.

Kozlovska, O. and F. Sadyrbaev, 2024 In search of chaos in genetic systems. Chaos Theory and Applications **6**: 13–18.

Leduc-Gaudet, J. P., S. N. Hussain, and G. Gouspillou, 2022 Parkin: a potential target to promote healthy ageing. The Journal of Physiology **600**: 3405–3421.

Löchel, H. F. and D. Heider, 2021 Chaos game representation and its applications in bioinformatics. Chaos **19**: 6263–6271.

Makova, K., B. Pickett, R. Harris, and et al., 2024 The complete sequence and comparative analysis of ape sex chromosomes. Nature **630**: 401–411.

Marín, I., J. I. Lucas, A. C. Gradilla, and A. Ferrús, 2004 Parkin and relatives: the rbr family of ubiquitin ligases. Physiological Genomics **17**: 253–263.

Miller, M. A., T. Schwartz, B. E. Pickett, S. He, E. B. Klem, *et al.*, 2015 A restful api for access to phylogenetic tools via the cipres science gateway. Evolutionary bioinformatics online **11**: 43–48.

Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, *et al.*, 2020 Iq-tree 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular biology and evolution **37**: 1530–1534.

Mitchell, M. and M. Newman, 2002 Complex systems theory and evolution. In *Encyclopedia of evolution*, edited by M. Pagel, pp. 1–5, Oxford University Press, New York, NY.

Mouchet, M. A. and D. Mouillot, 2010 Decomposing phylogenetic entropy into , and components. Biology Letters **6**: 665–668.

Munk, S. H., V. Voutsinos, and V. H. Oestergaard, 2021 Large intronic deletion of the fragile site gene prkn dramatically lowers its fragility without impacting gene expression. Frontiers in Genetics **12**.

Olszewska, D. A., A. McCarthy, A. I. Soto-Beasley, *et al.*, 2022 Parkin, pink1, and dj1 analysis in early-onset parkinson's disease in ireland. Irish Journal of Medical Science **191**: 901–907.

Patidar, V. and G. Kaur, 2024 Lossless image encryption using robust chaos-based dynamic dna coding, xoring and complementing. Chaos Theory and Applications **6**: 13–18.

Perez-Gala, R., Y. S. Cangrejo-Useda, C. V. Castillo-Lopez, M. J. Sanchez-Manjarres, C. E. Cadena-Caballero, *et al.*, 2025 Frequency chaos game method and fractals show evolutionary relationships of the prkn gene in primates.

Retnaningsih, R., 2024 Fractal geometry, fibonacci numbers, golden ratios, and pascal triangles as designs. The Journal of Academic Science **1**: 51–66.

Rodrigo-Gala, 2025 Complexity-science-group-chaos-fractals-nature-aplications. GitHub repository.

Saeed, M., 2020 Fractal genomics of sod1 evolution. Immunogenetics **72**: 439–445.

Siegenfeld, A. F. and Y. Bar-Yam, 2020 An introduction to complex systems science and its applications. Complexity **2020**: 6105872.

Sievers, A., L. Sauer, M. Hausmann, and G. Hildenbrand, 2021

Eukaryotic genomes show strong evolutionary conservation of k-mer composition and correlation contributions between introns and intergenic regions. Genes **12**.
Tanaka, K., 2020 The pink1-parkin axis: An overview. Neuroscience Research **159**: 9–15.
Vischer, E. and E. Chargaff, 1948 The composition of the pentose nucleic acids of yeast and pancreas. Journal of Biological Chemistry **176**: 715–734.
Wang, X. S., T. R. Cotton, S. J. Trevelyan, *et al.*, 2023 The unifying catalytic mechanism of the ring-between-ring e3 ubiquitin ligase family. Nature Communications **14**: 14.

How to cite this article: Perez-Gala, R., Cangrejo-Useda, Y. S., Castillo-Lopez, C. V., Sanchez-Manjarres, M., Cadena-Caballero, C. E., Martinez-Perez, F., and Ramirez-Alatriste, F. Frequency chaos game method and fractals show evolutionary relationships of the *PRKN* gene in primates. *Chaos Theory and Applications*, x(x), xxx-xxx, 2025.

Table 1 Programs

Software Name	Function
Code Random DNA generator (RandDNA)	From a DNA sequence, measure its length and generate a randomly shuffled sequence of the same size. Check for nucleotides that match the original sequence and randomly change them.
Code DNA Subsequence Forward (DSF)	Take a base DNA sequence and a primary DNA sequence. Measure the length of the primary sequence and extract a subsequence from the base sequence that matches this length, starting from the first nucleotide of the base sequence.
Code DNA Subsequence Reverse (DSR)	Take a base DNA sequence and a primary DNA sequence. Measure the length of the primary sequence and extract a subsequence from the base sequence that matches this length, starting from the last nucleotide and moving backwards.
Code DNA Extraction Random (DNAxtra)	Take a base DNA sequence and a primary DNA sequence. Measure the length of the primary sequence and extract a subsequence from the base sequence that matches this length, starting at a random point in the base sequence. Repeat this extraction "n" times.
Code Random Mutation Entropy Analysis (RMEA)	Select DNA fragments and make "n" random modifications to the nucleotides of each fragment. This process is repeated "x" times for the same sequence, calculating Shannon Entropy for each case and obtaining the average. The results are recorded in an Excel file and presented in a graph.
Code Complementary Nucleotide Sequence Mutation (CNSMut)	Extract DNA subsequences and perform "n" nucleotide replacements, substituting them with their complementary counterparts. This procedure is applied "x" times to each sequence, determining the average Shannon Entropy. The data is stored in an Excel file and accompanied by a graph.
Code Chemical Swap Entropy Analysis (ChemSwap)	Obtain DNA fragments and replace "n" nucleotides in each fragment with those belonging to the opposite chemical group. Repeat this procedure "x" times for the same sequence, calculate the Shannon Entropy for each iteration, and average the obtained values. The results are saved in an Excel file and visualized in a graph.
Code IMAGE GROUPER (IMG)	Take a series of images and combine them into a new image.