



FRACTALES EN GENOMAS: TRANSICIONES DE FASE, DE PROCARIONTES A EUCARIONTES

Roberto Álvarez

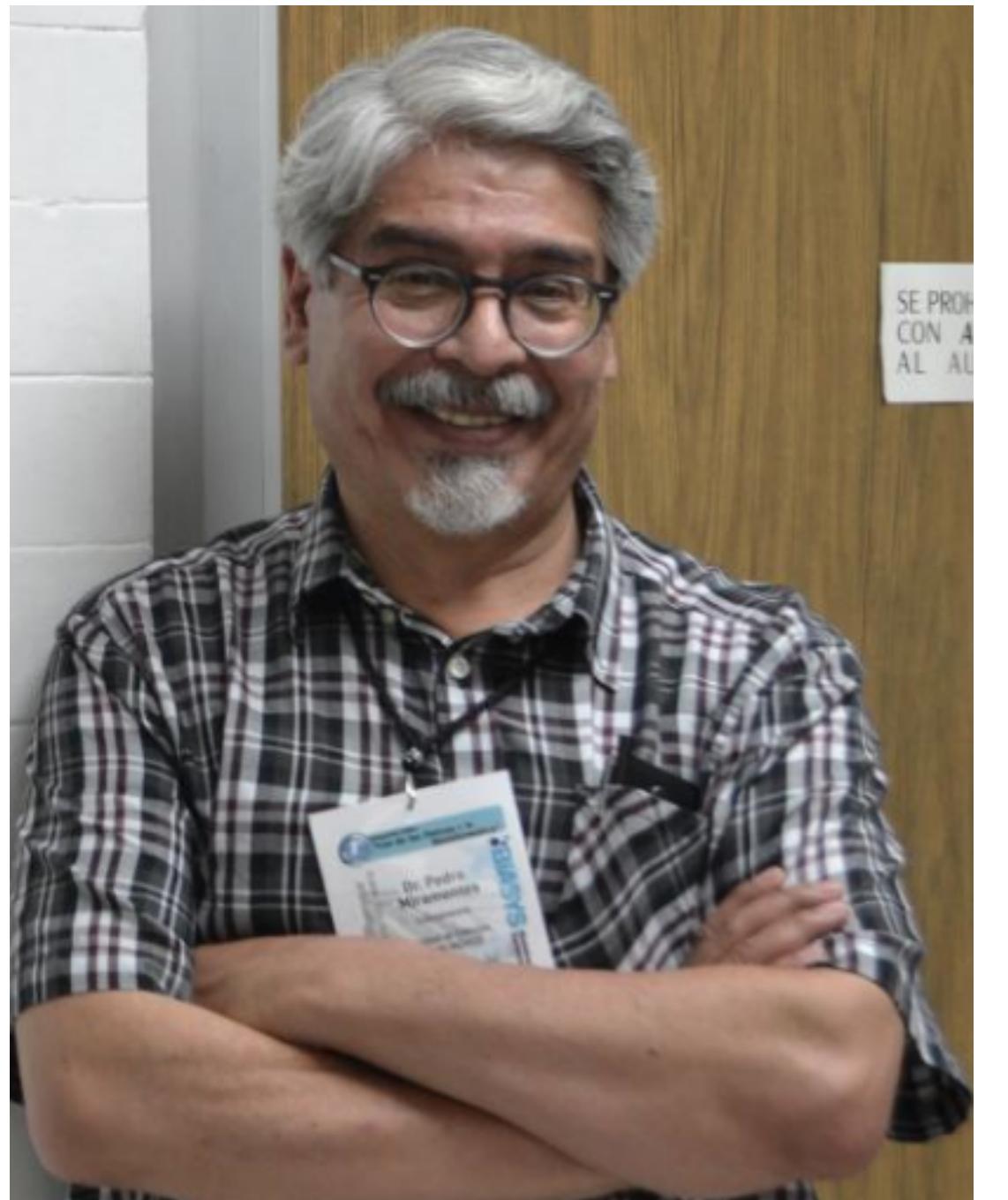
Universidad Autónoma de Querétaro

¿QUÉ DISCUTIREMOS?

ESTRUCTURA

- ▶ El juego del caos sobre secuencias de ADN
 - ▶ Patrones fractales
- ▶ De CGR a FGR
- ▶ Formalismo termodinámico
- ▶ Transiciones de fase
 - ▶ Mapeos unidimensionales

POR SI ME QUEDO SIN TIEMPO



Pedro Miramontes

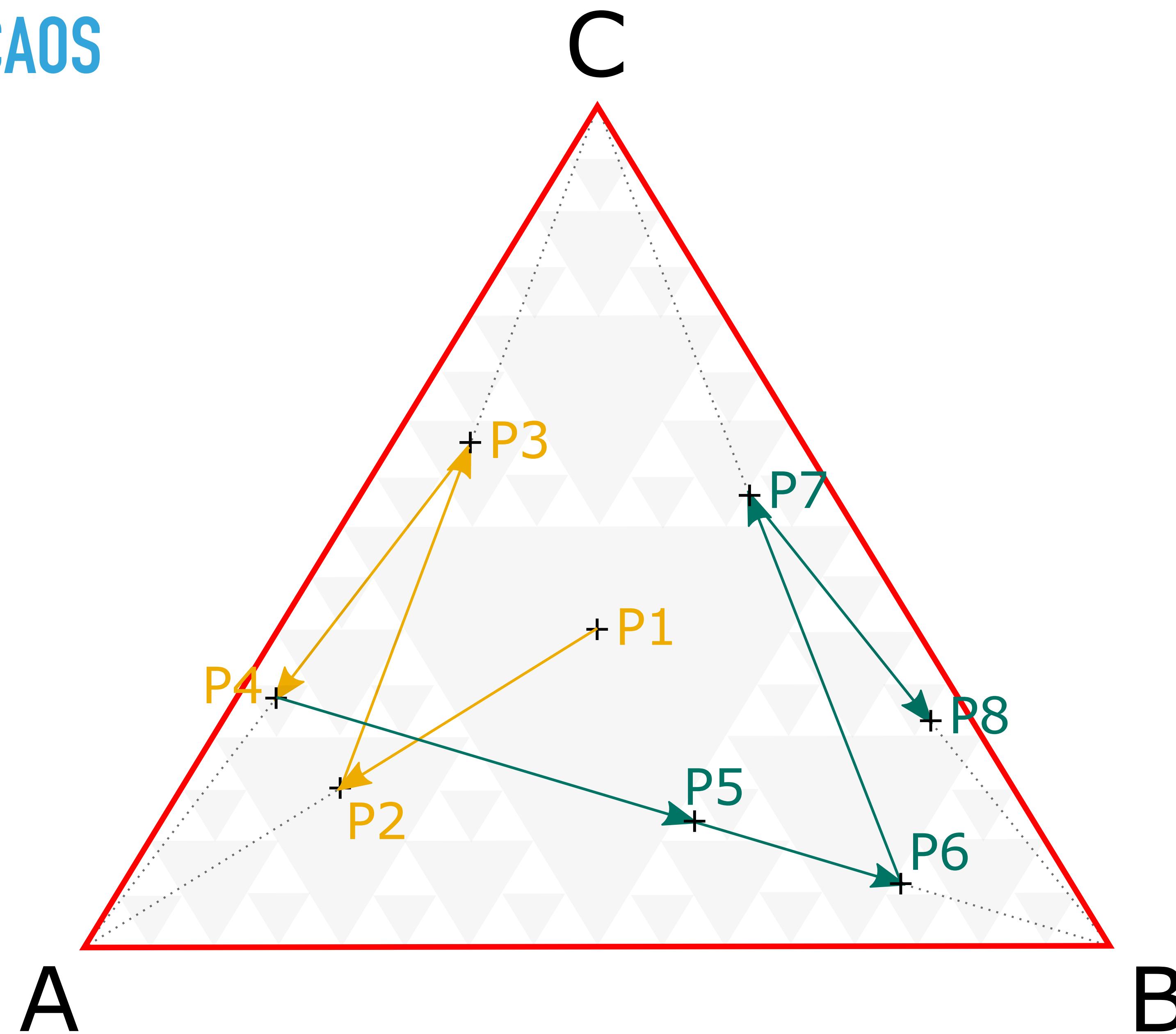
FC-UNAM



Enrique Hernández Lemus

INMEGEN

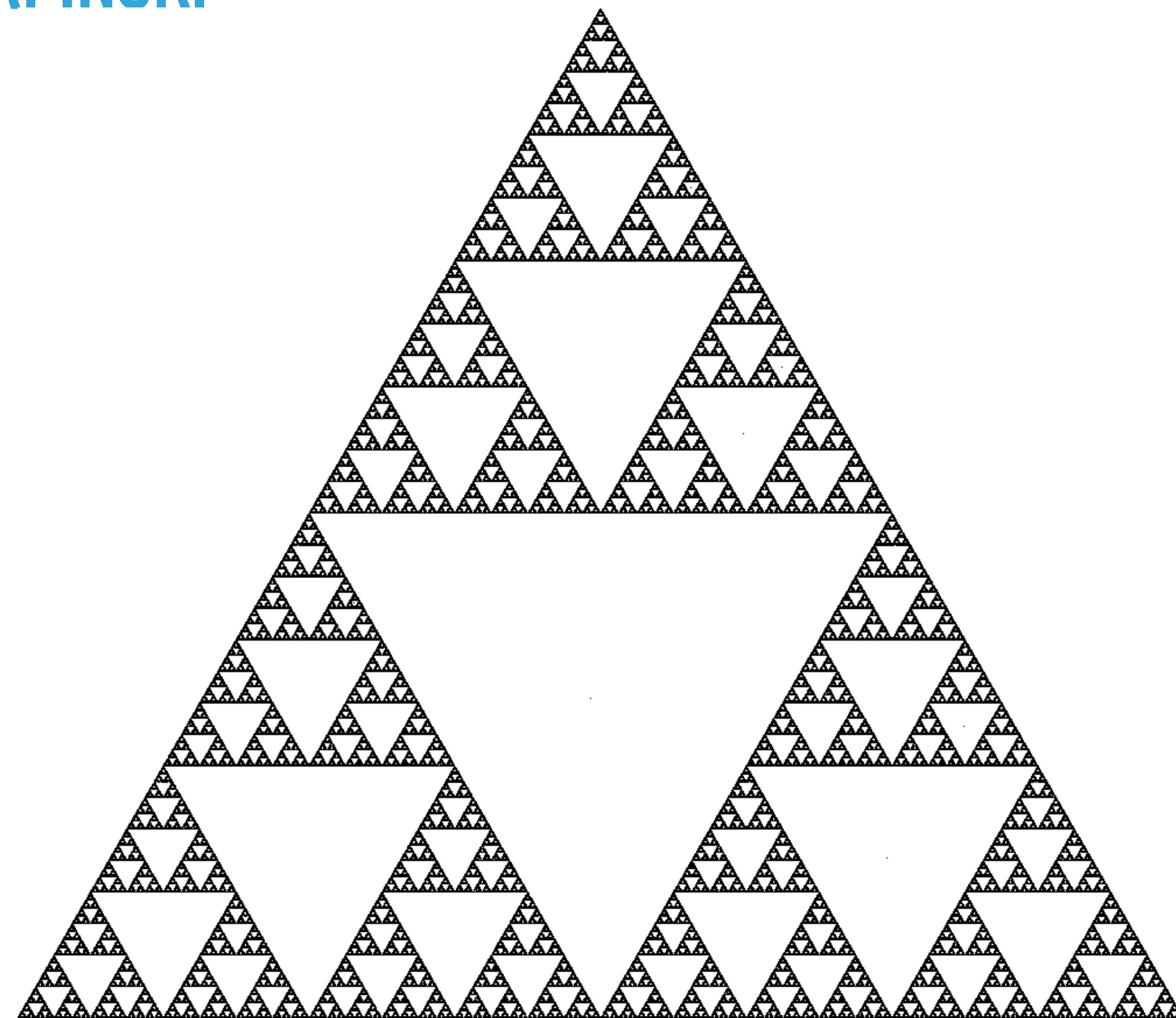
EL JUEGO DEL CAOS



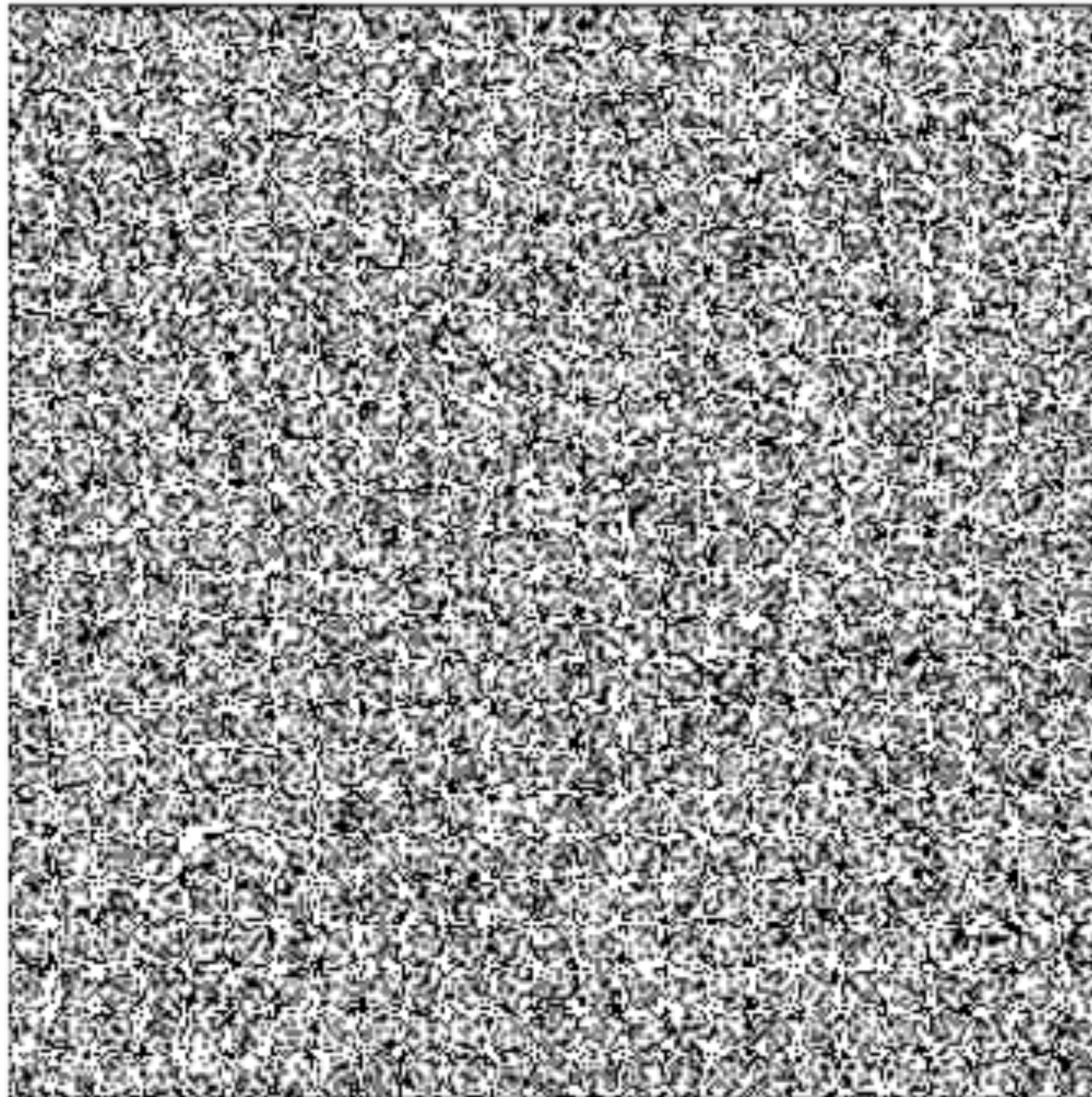
TRIÁNGULO DE SIERPINSKI

$$n = 3$$

$$r = \frac{1}{2}$$



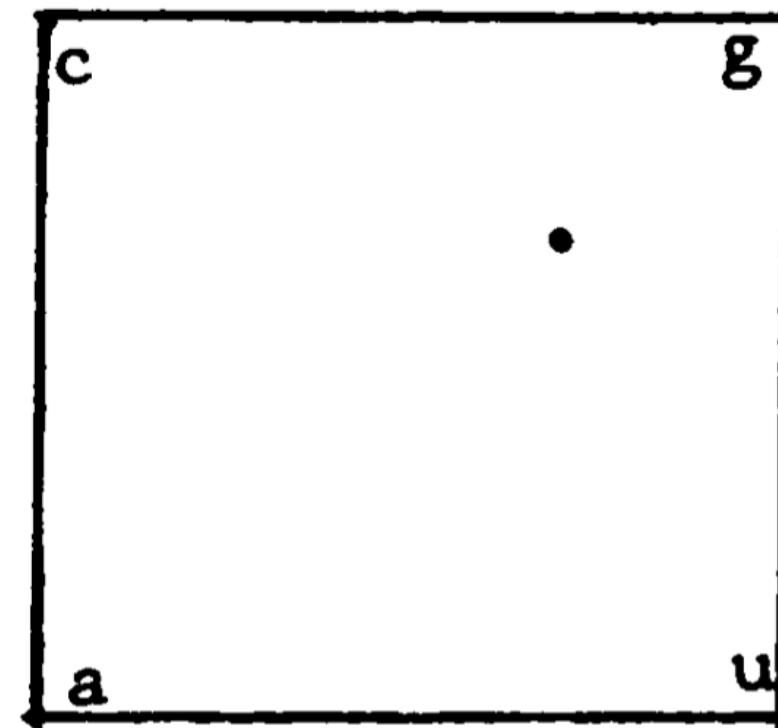
CUADRADO CON DENSIDAD UNIFORME



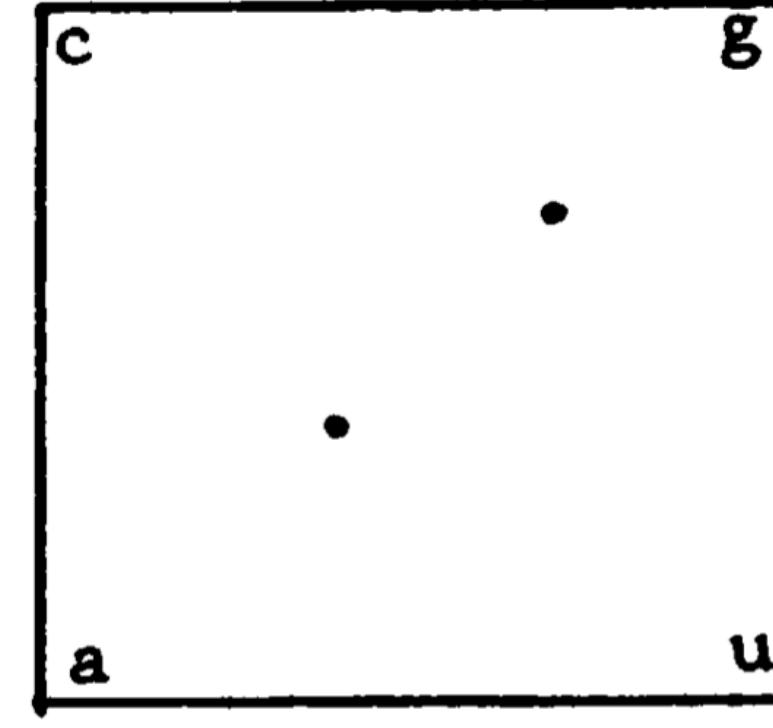
$$n = 4$$

$$r = \frac{1}{2}$$

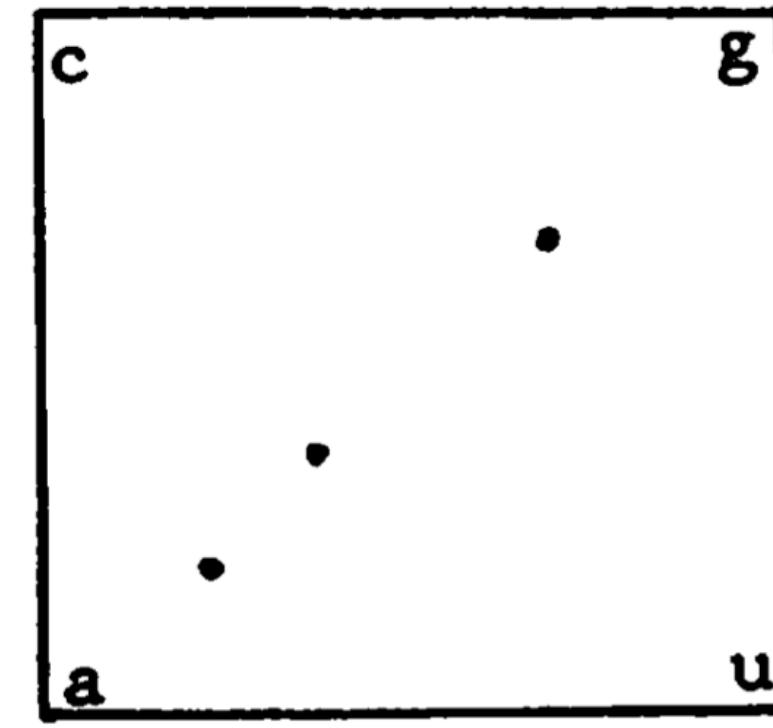
EL JUEGO DEL CAOS CON ADN



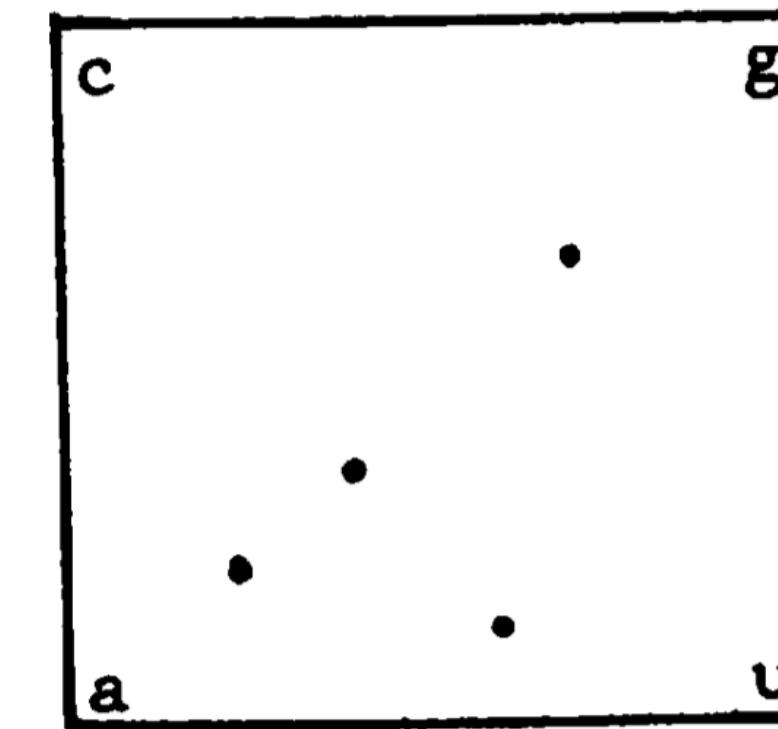
"g"



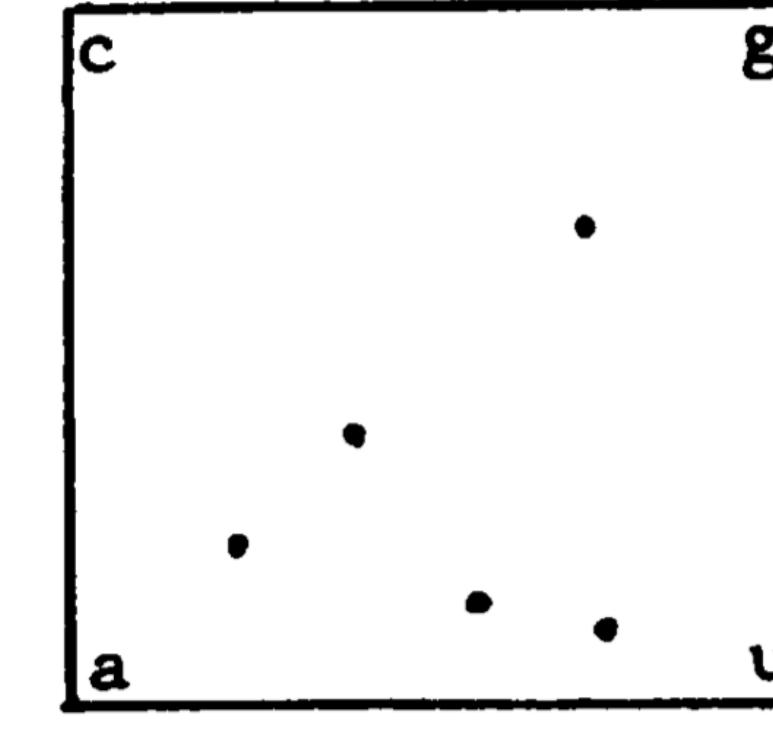
"ga"



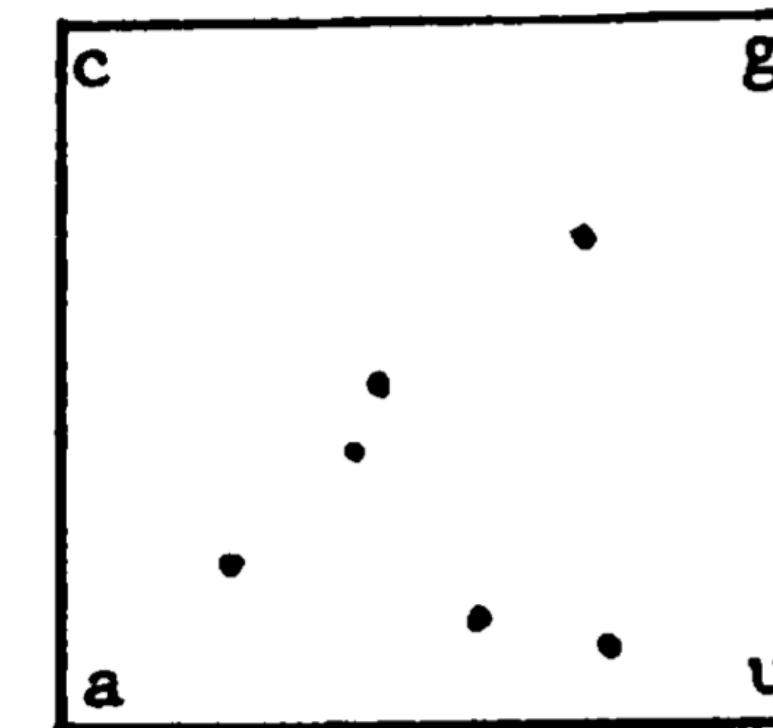
"gaa"



"gaat"

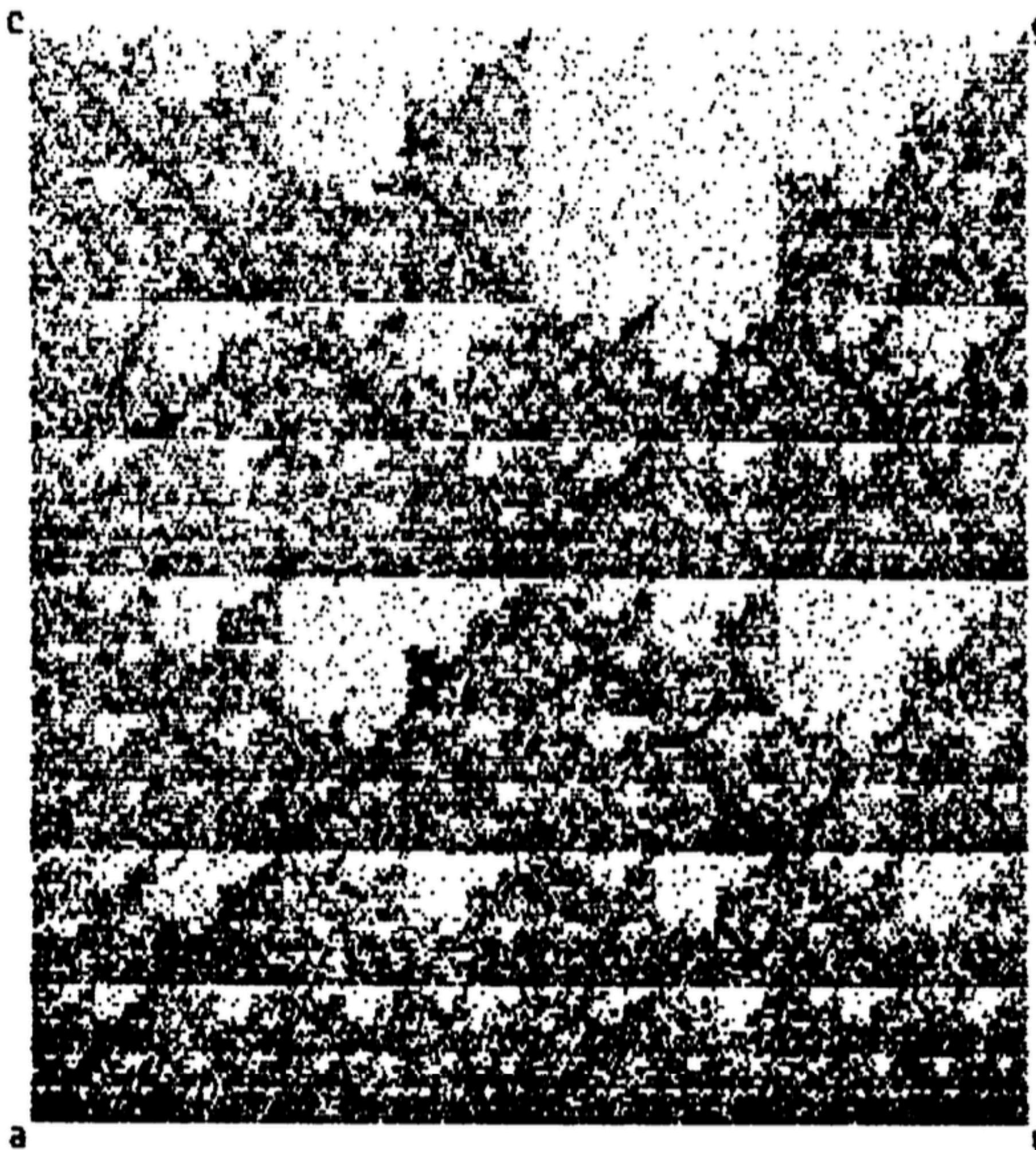


"gaatt"



"gaattc"

LOS PATRONES NO SON ALEATORIOS EN EL ADN



Human Beta Globin 7 Kb

CHAOS GAME REPRESENTATION (CGR)

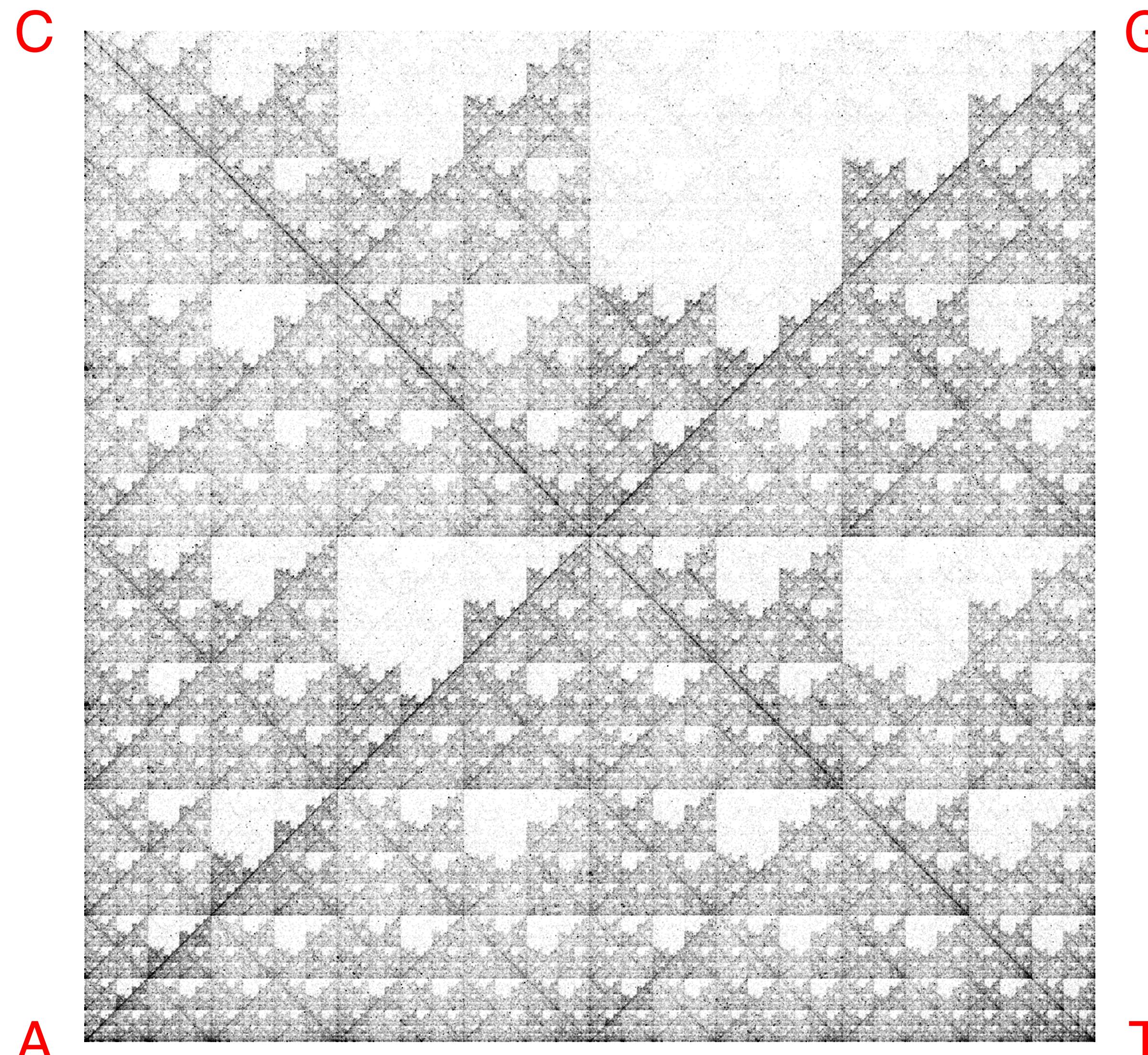
- ▶ Una secuencia se representa como un patrón único.
- ▶ Una secuencia se asigna a a coordenadas únicas
- ▶ Un CGR mapea todas las secuencias posibles de todas las longitudes posibles a un espacio 2D o 3D.
- ▶ Una sola coordenada codifica la entrada de secuencia completa.
- ▶ El punto de partida apenas influye en el resultado.

¿QUÉ SE PUEDE HACER CON ESO?

USOS DEL CGR

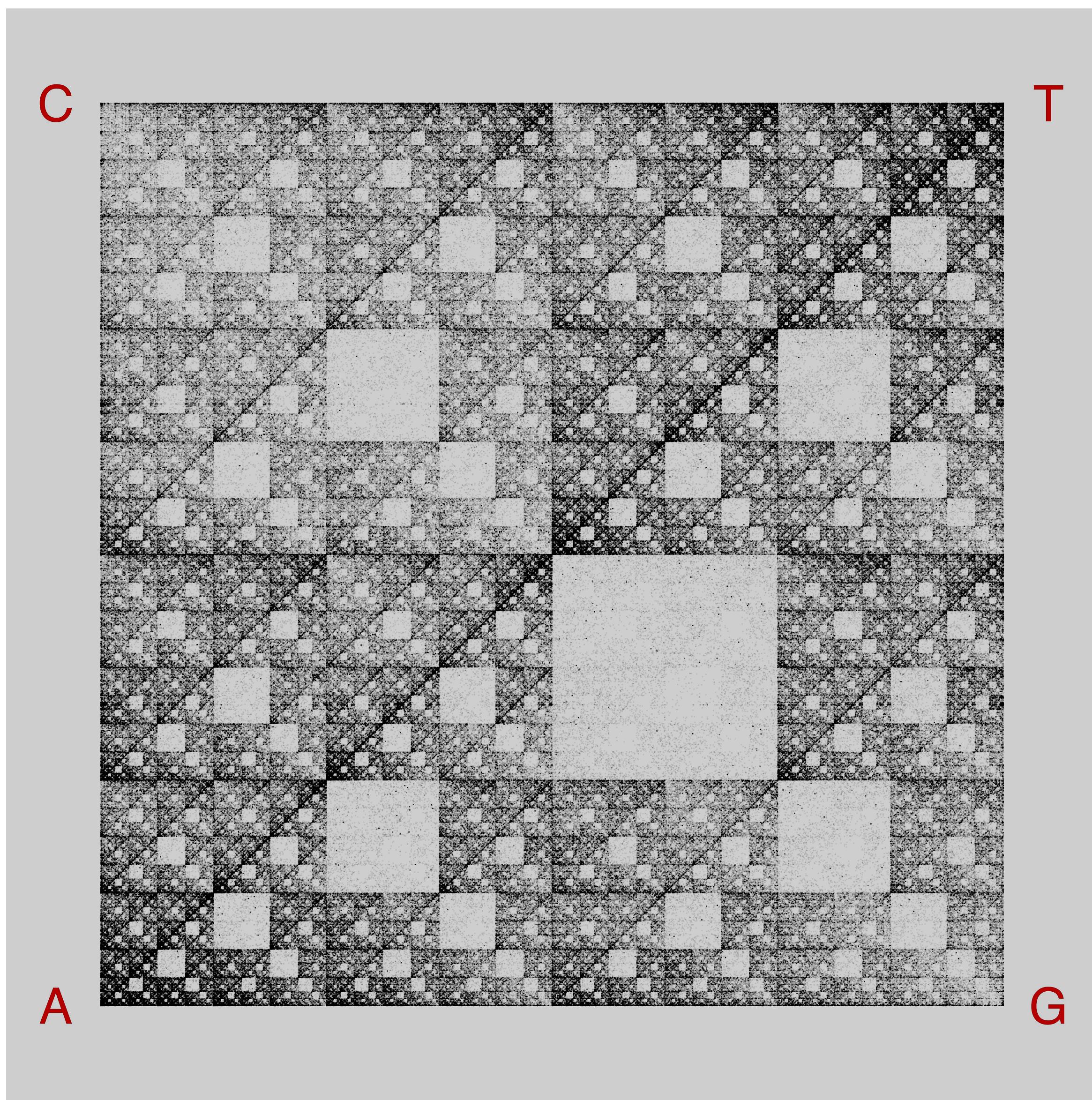
- ▶ Comparación de secuencias para la construcción de filogenias sin necesidad de alineamiento.
- ▶ Identificar patrones de enfermedades mitocondriales
- ▶ Insumo para algoritmos de inteligencia artificial

CROMOSOMA X



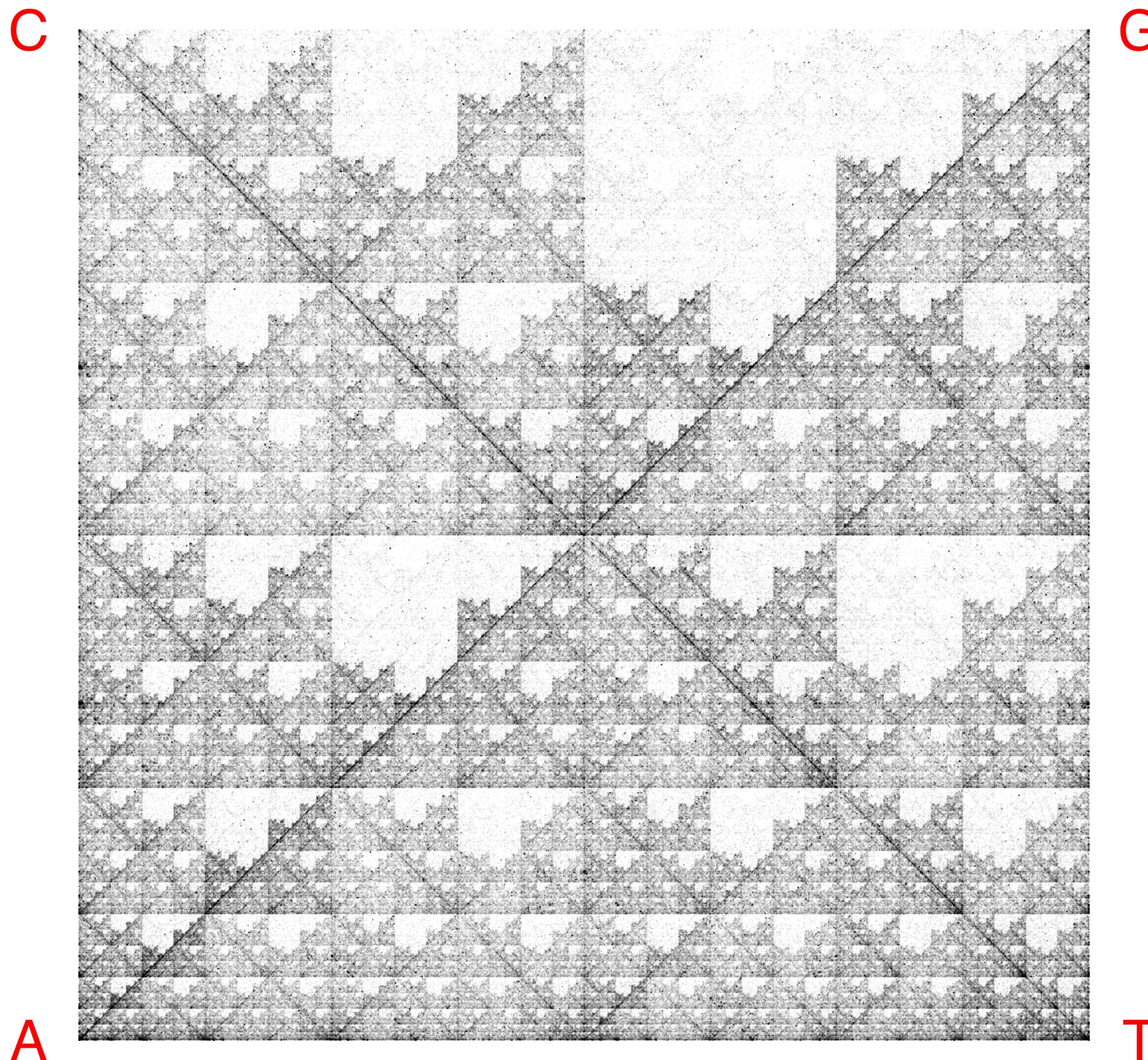
146 Mb

FRACTALES

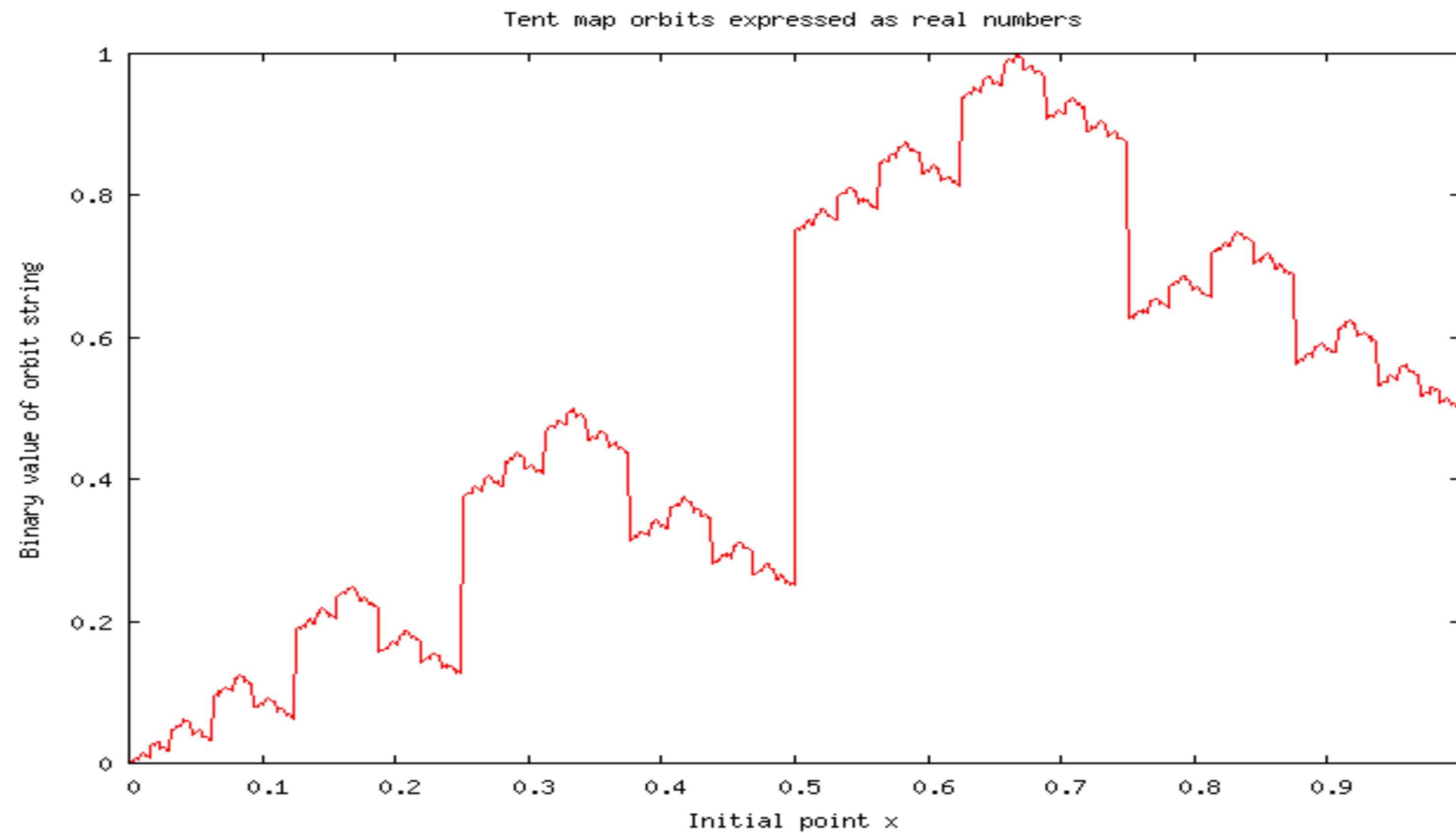


194 Mb

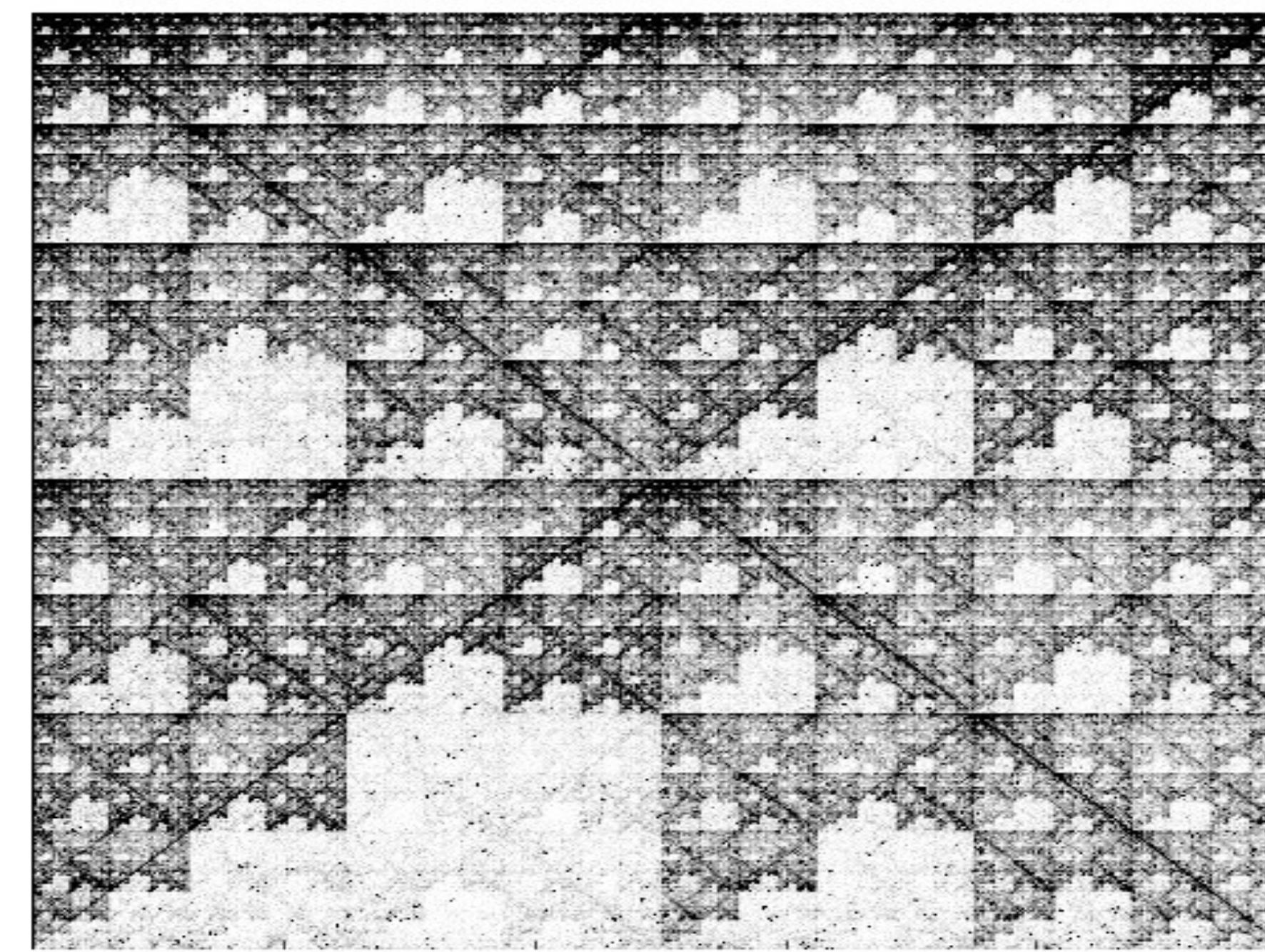
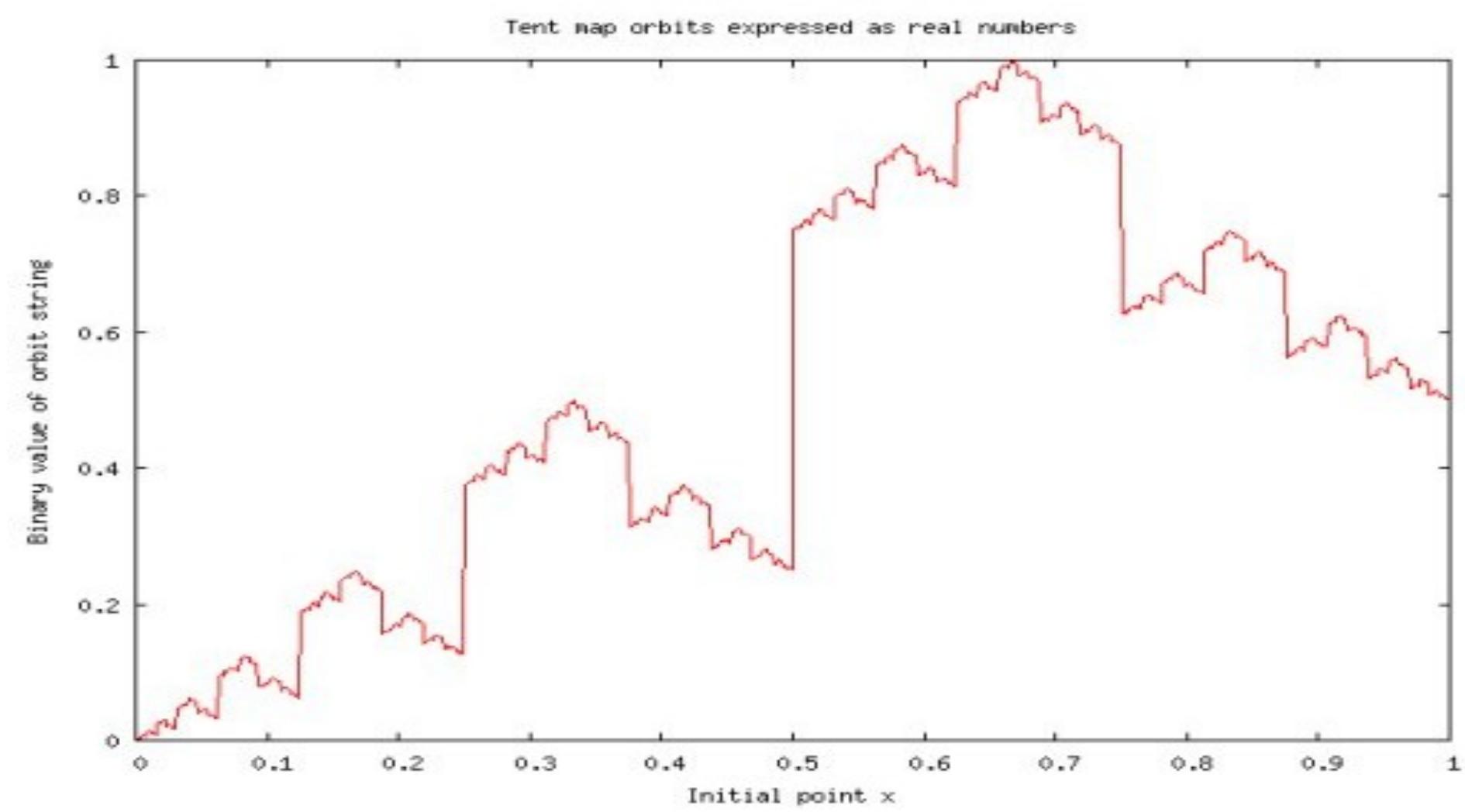
FRACTALES



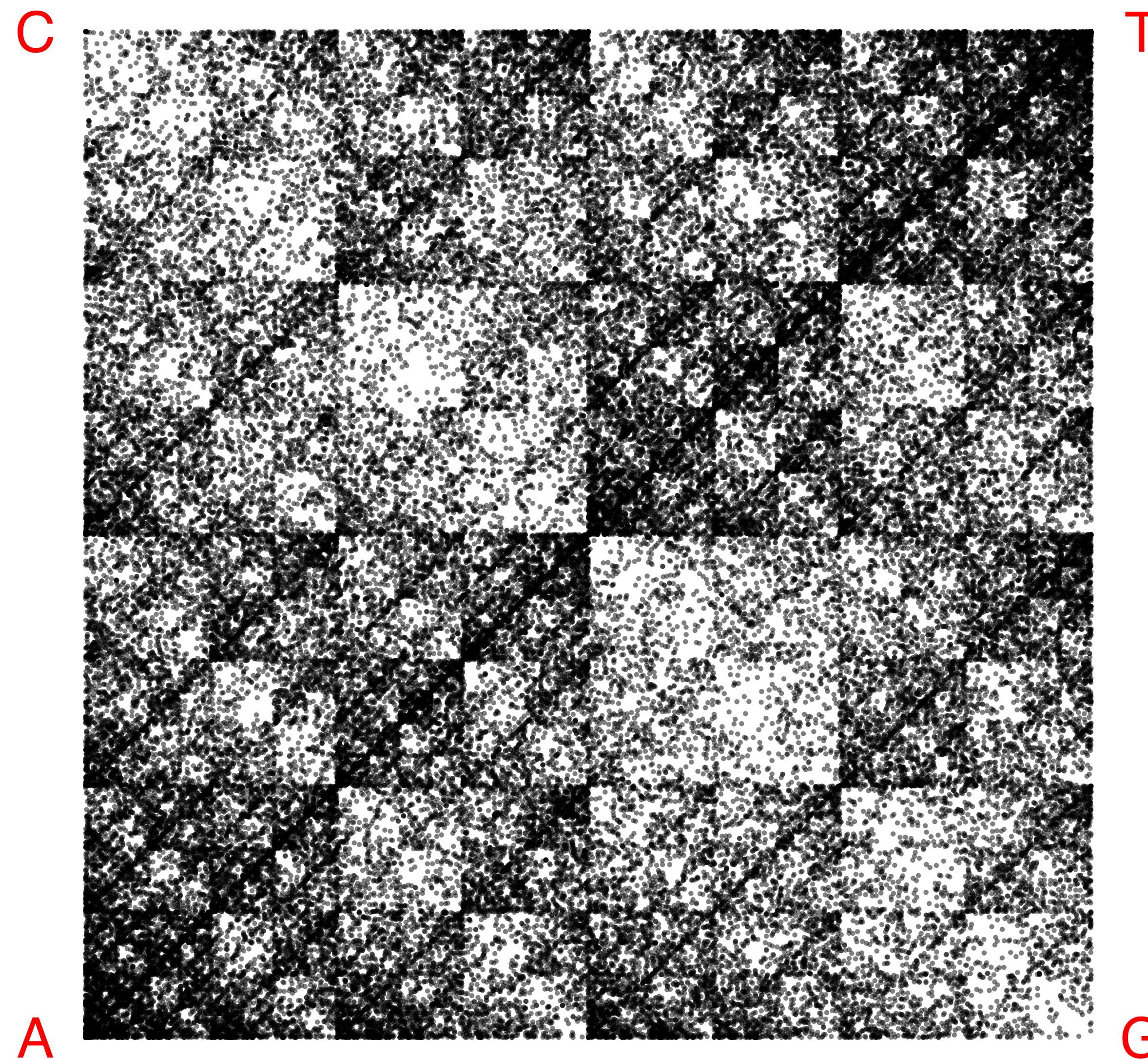
FRACTAL MATEMÁTICO



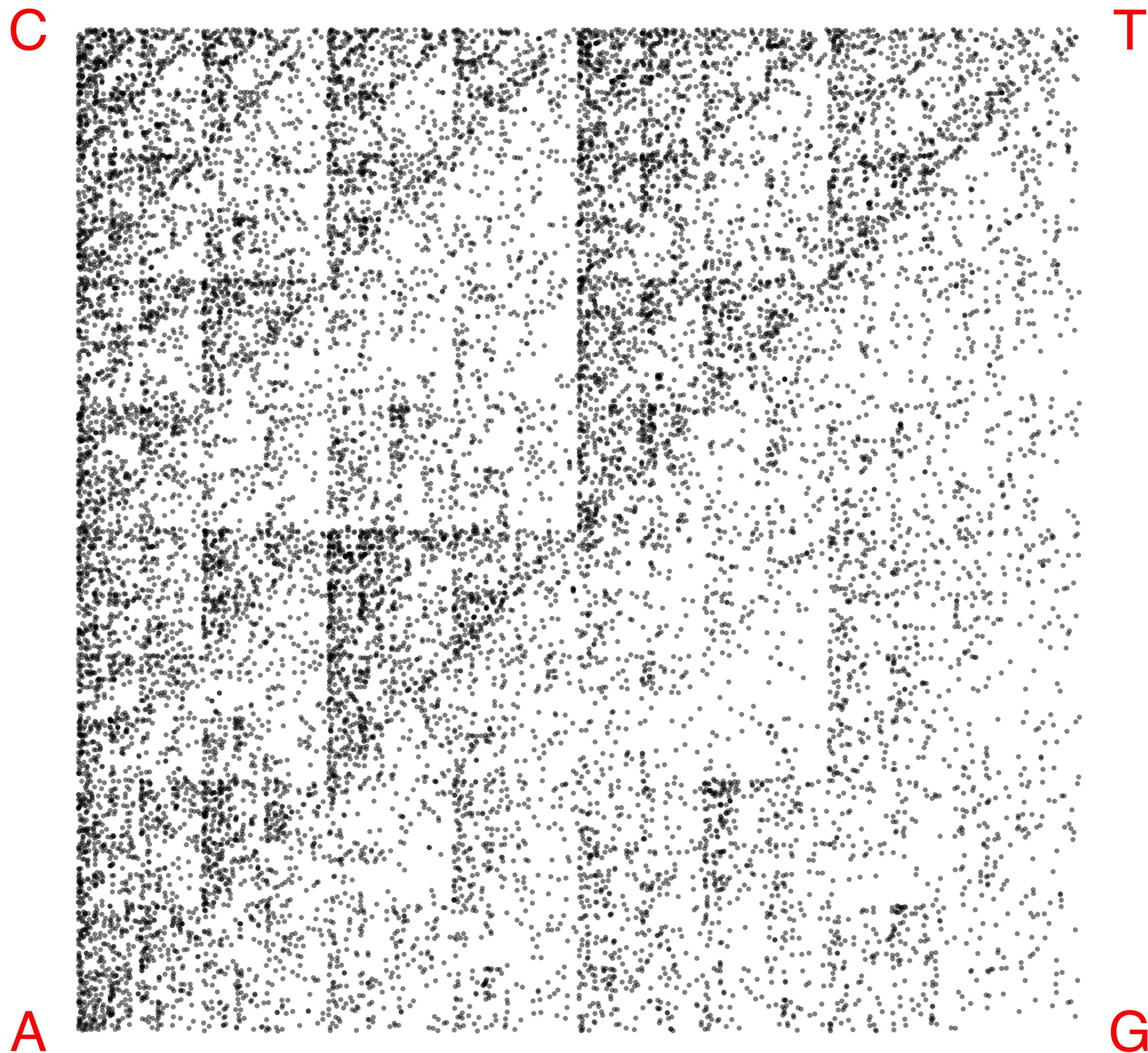
¿CÓMO EXPLICAMOS ESTO?



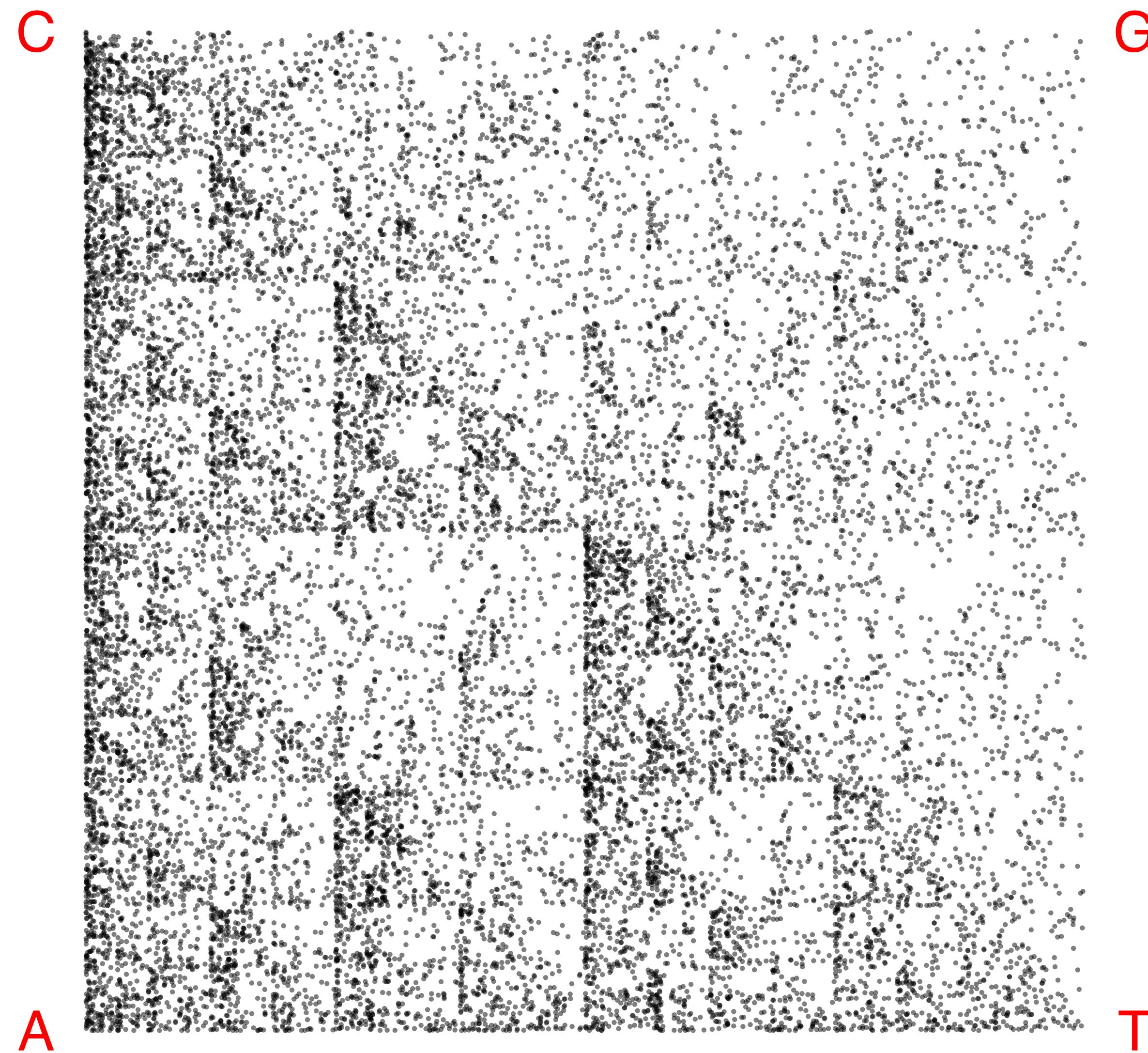
ARABIDOPSIS THALIANA



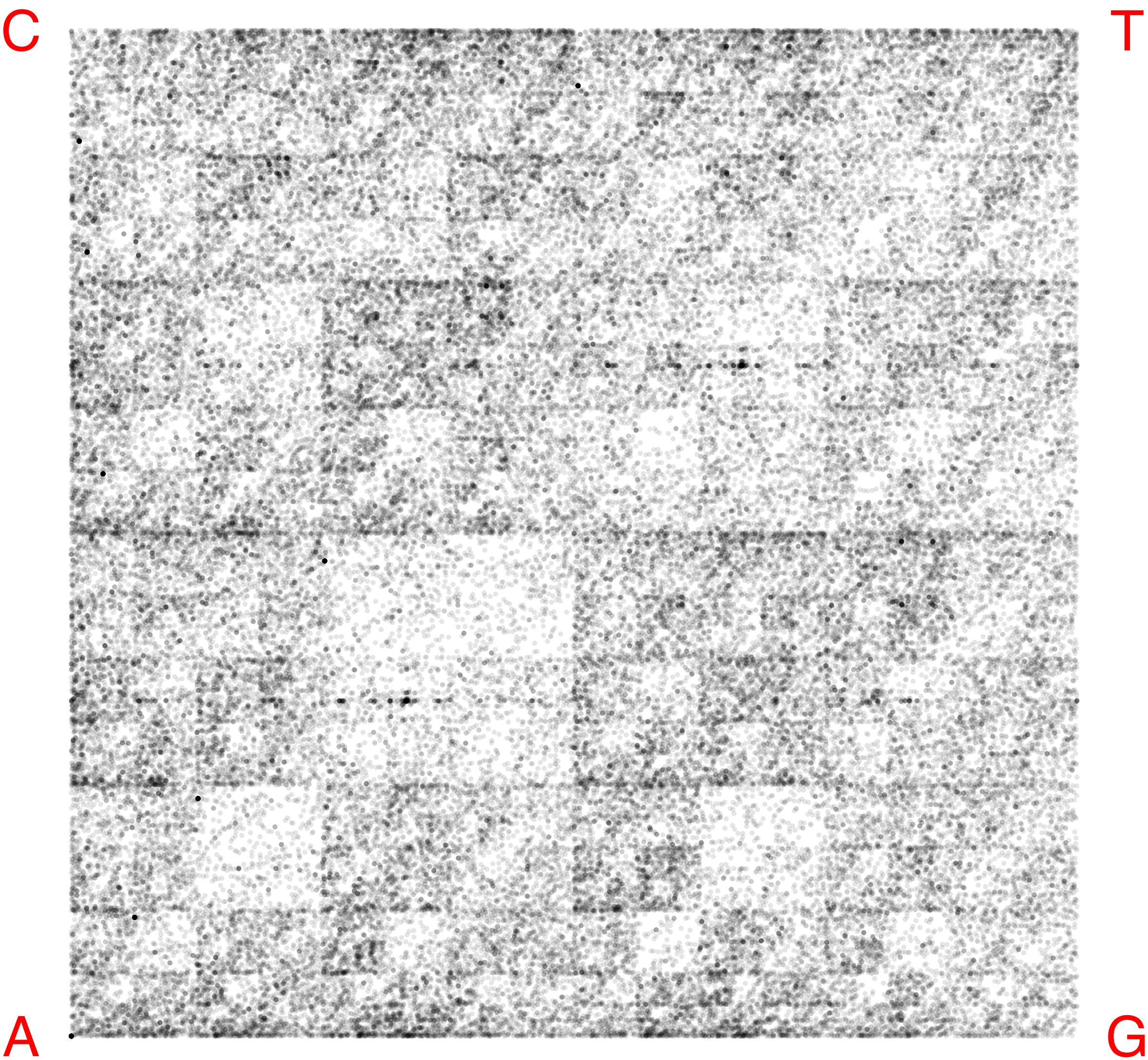
GENOMAS MITOCONDRIALES



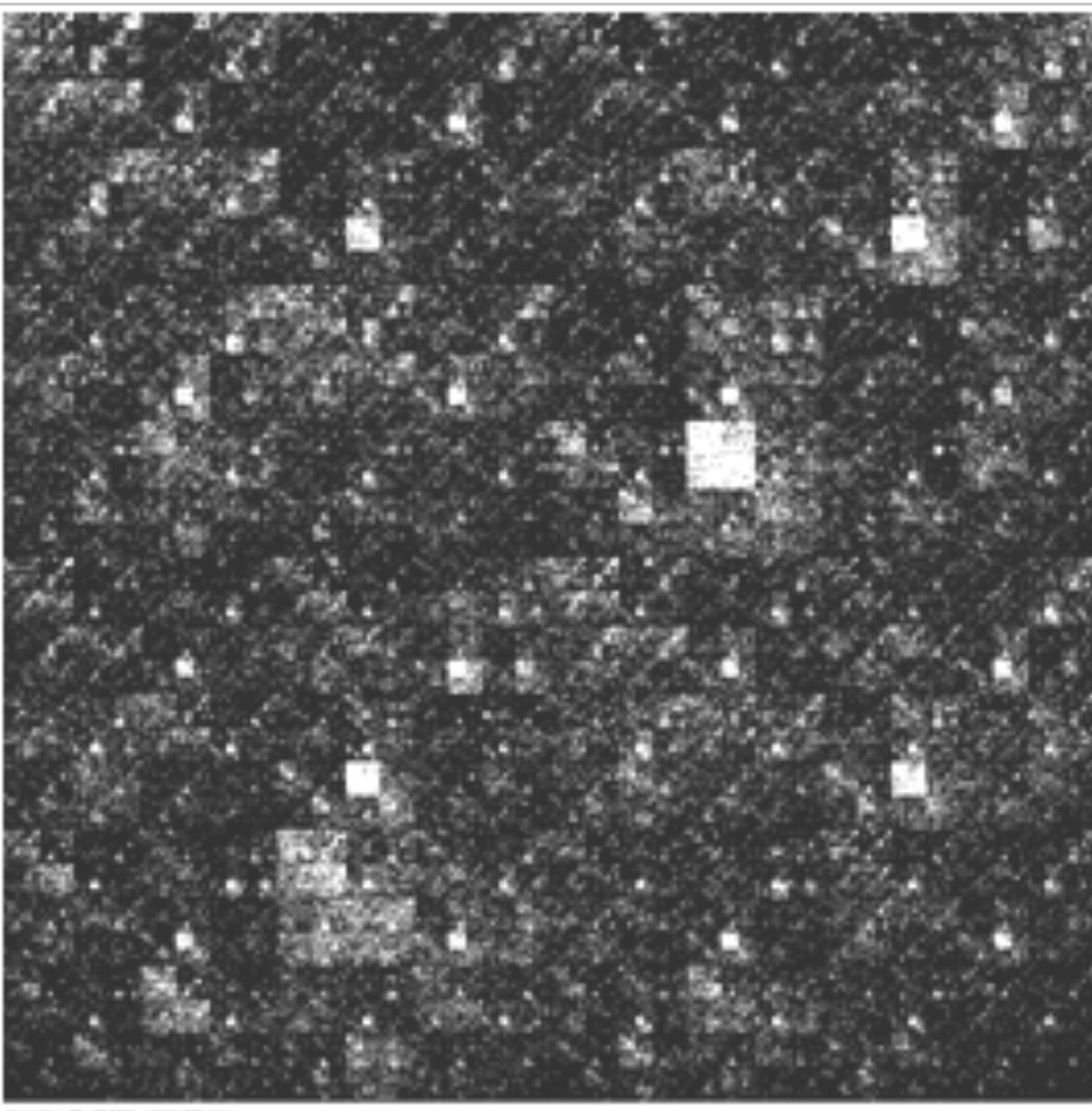
GENOMA MITOCONDRIAL



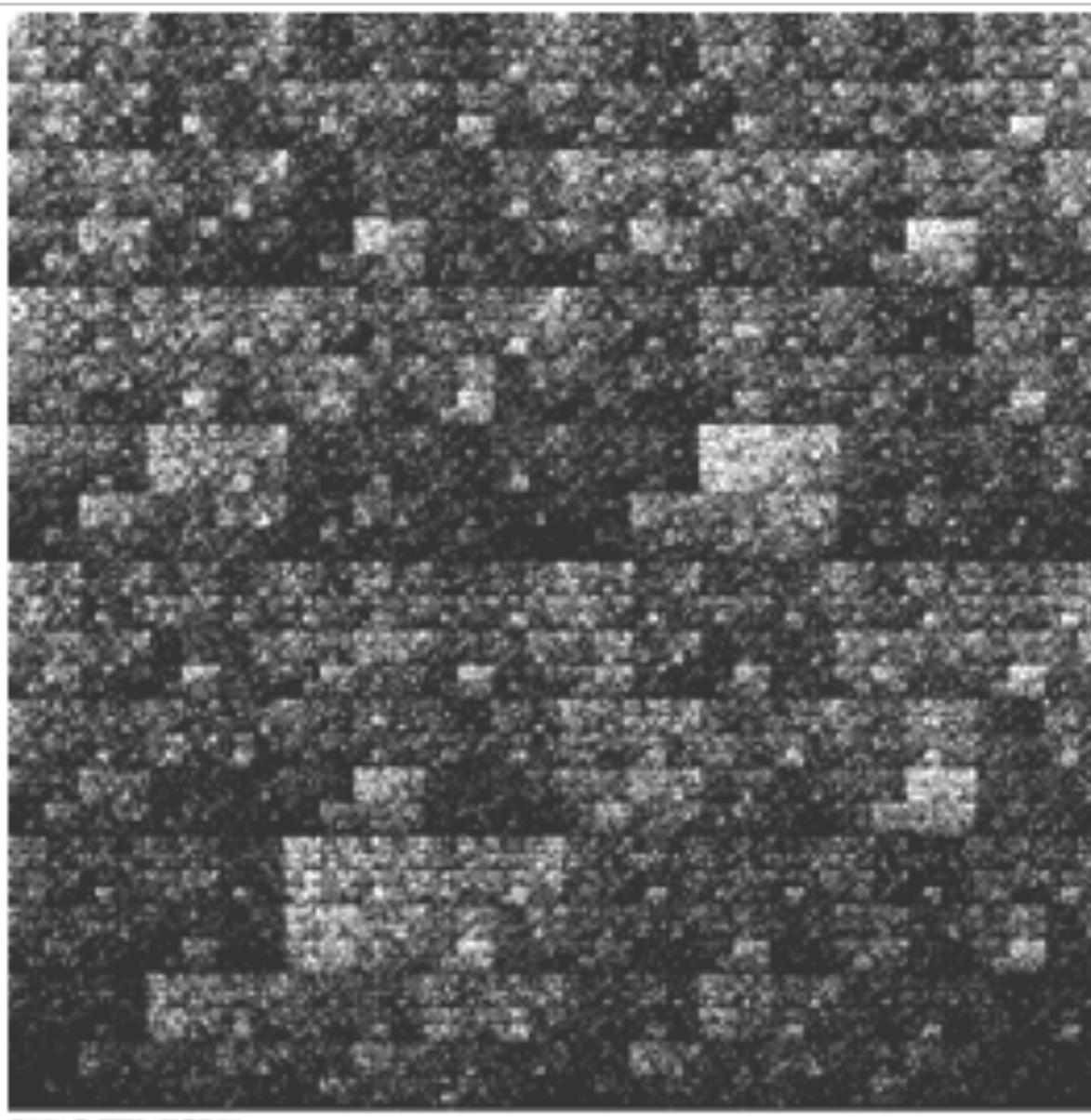
TOXOPLASMA GONDII CROMOSOMA 3



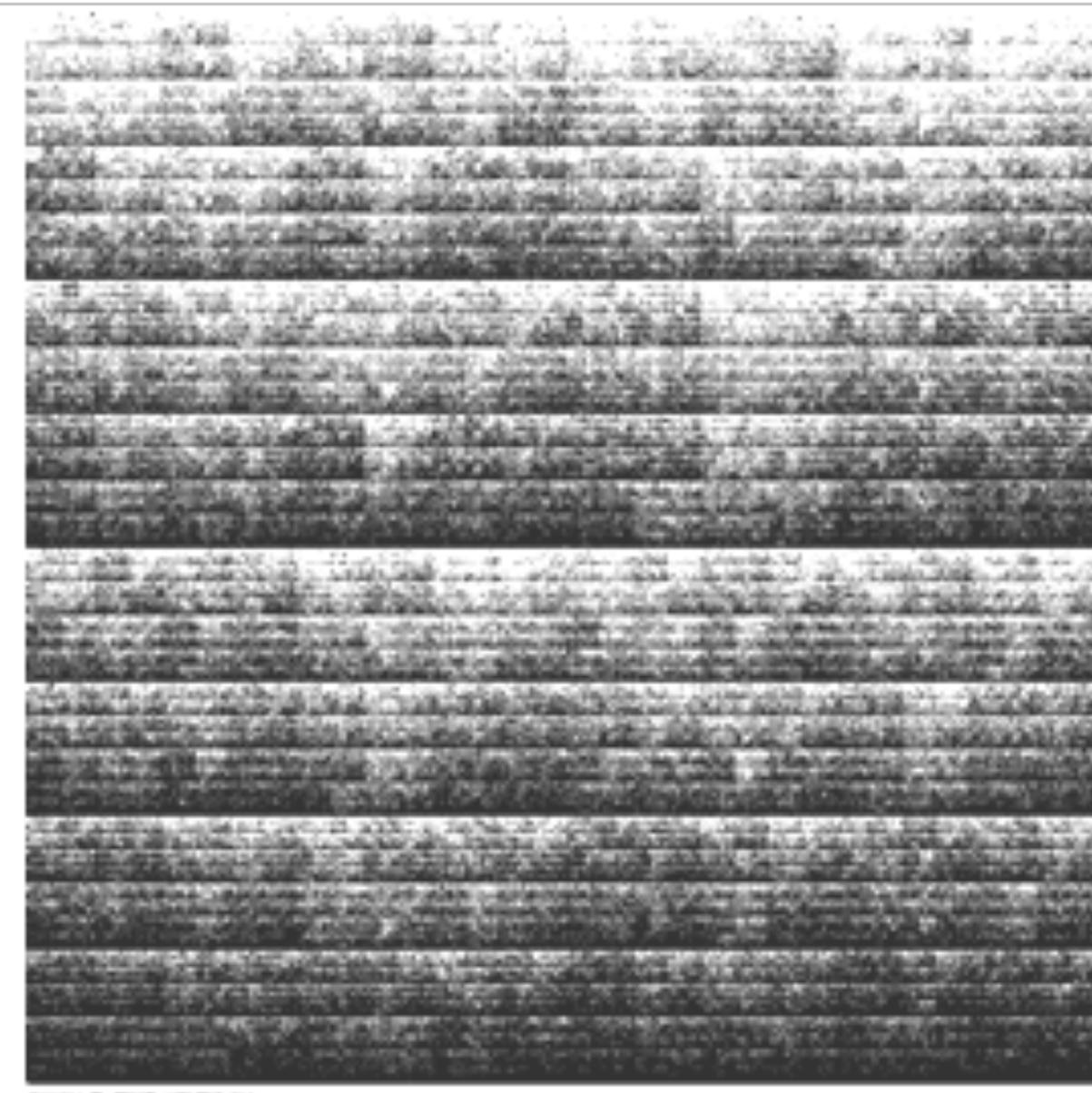
¿SIEMPRE SON FRACTALES?



E. coli

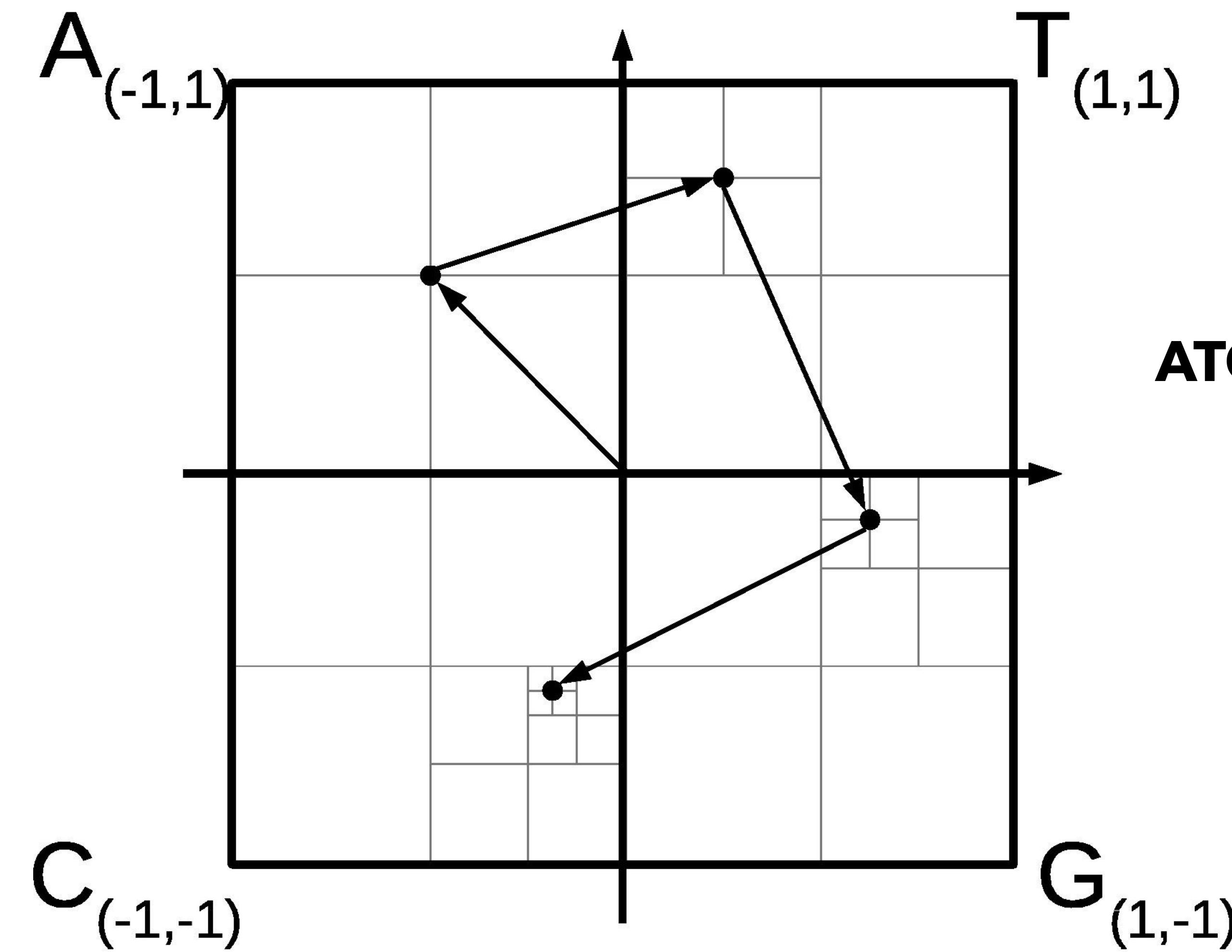


B. cereus

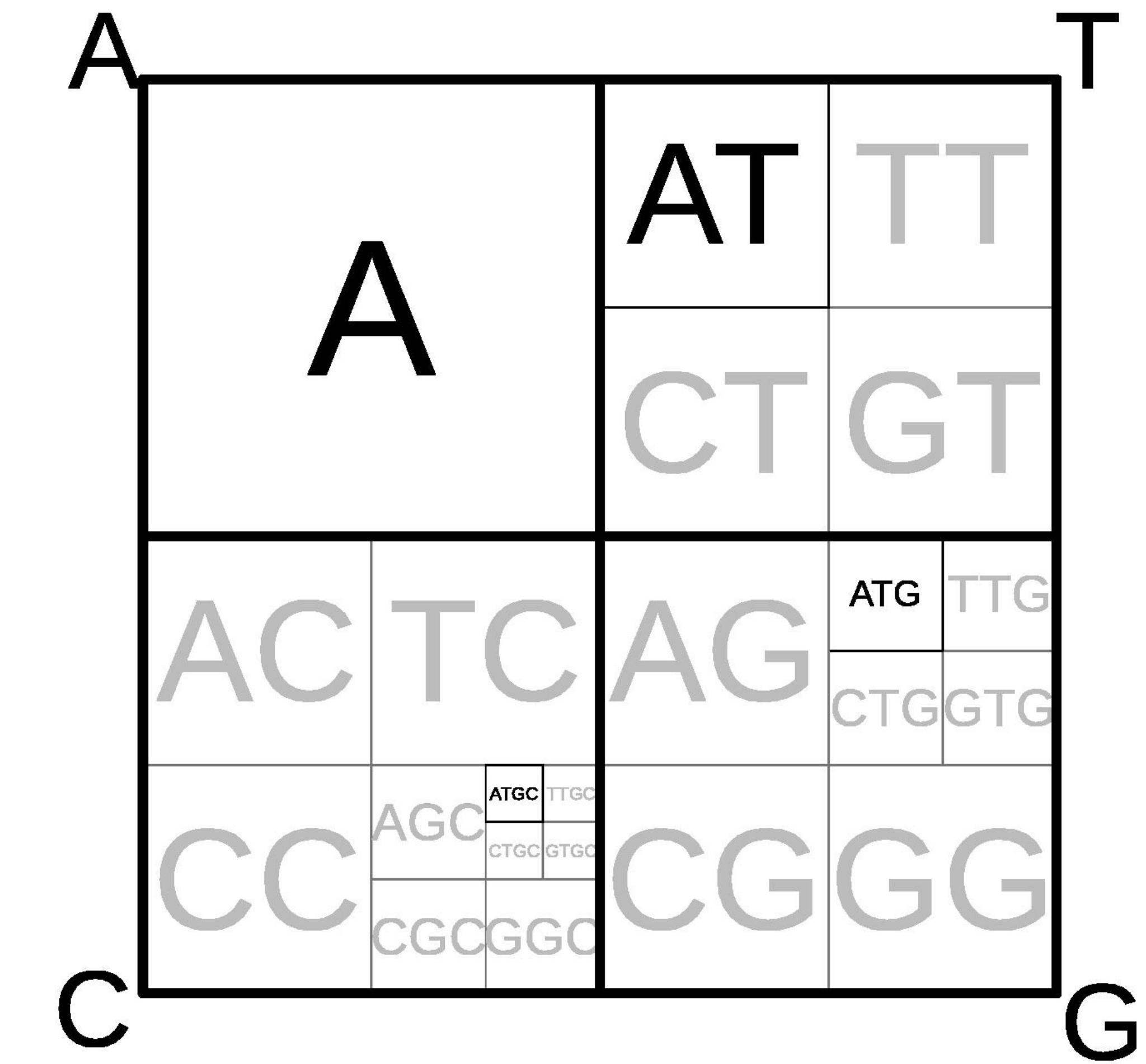


S. aureus

¿CÓMO EXPLICAMOS ESO?

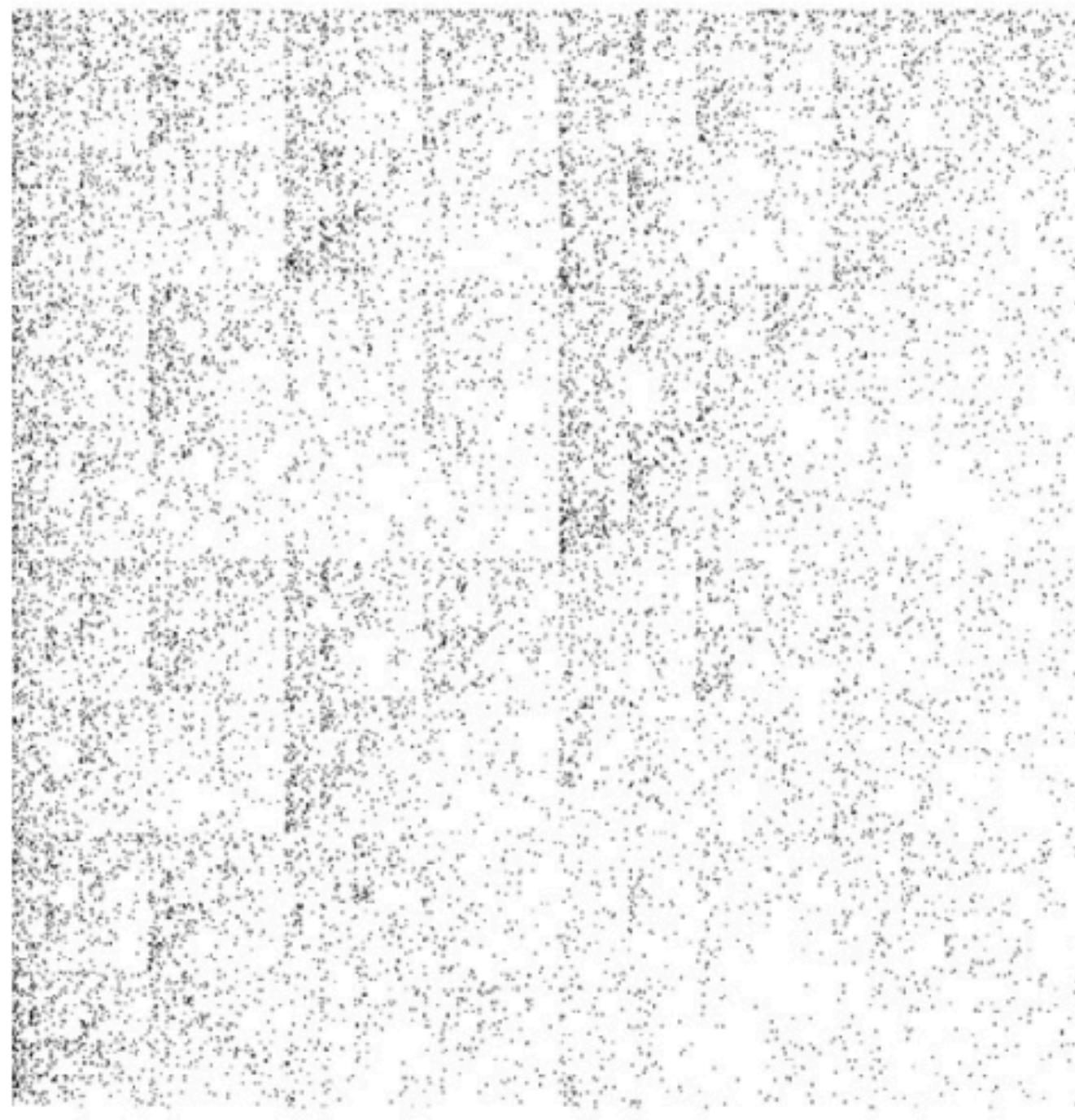


ATGC

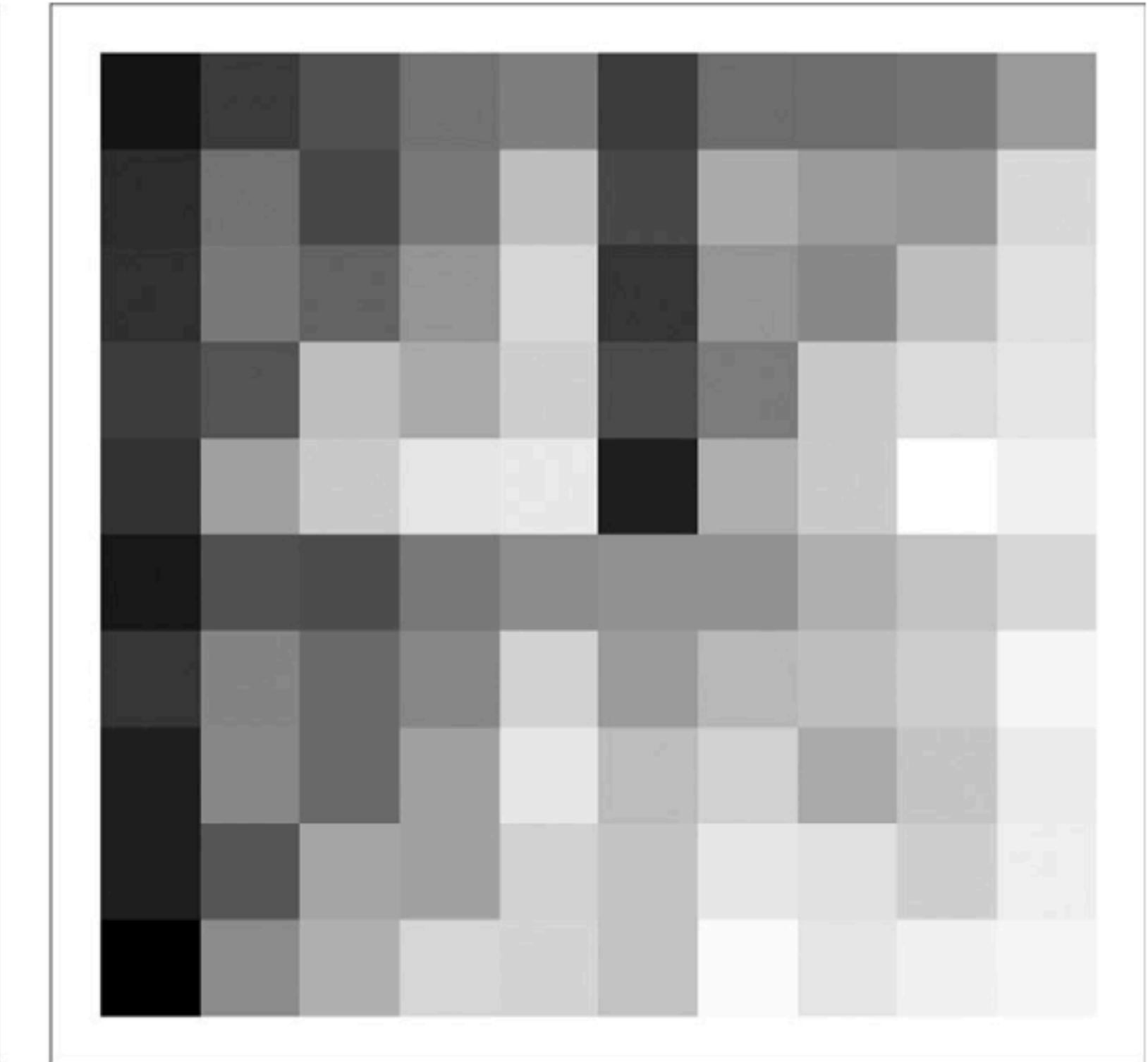


DENSIDAD DE PUNTOS

DE CGR A FREQUENCY CHAOS REPRESENTATION (FCR)

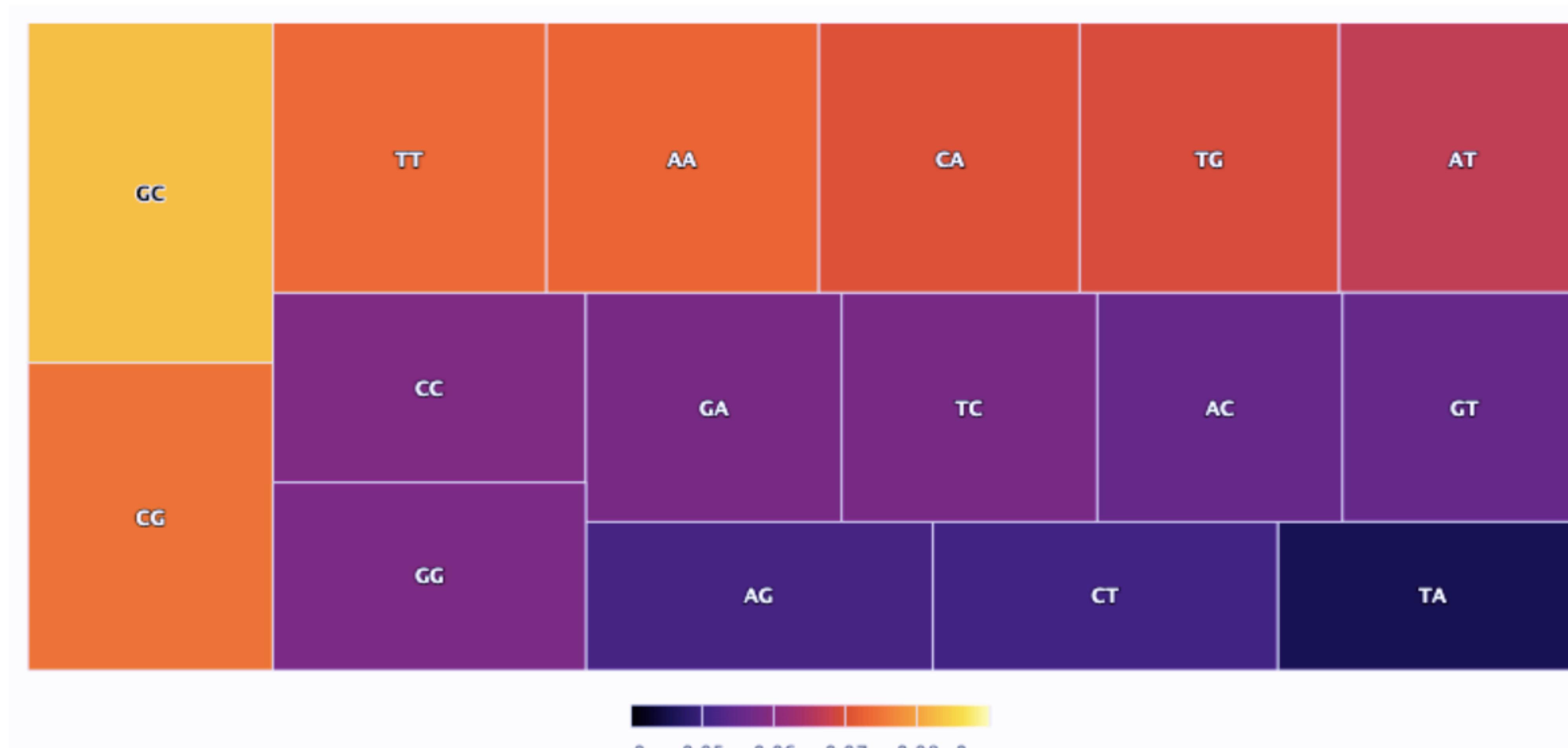


| | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|
| 519 | 411 | 364 | 329 | 377 | 320 | 331 | 253 |
| 464 | 201 | 522 | 152 | 416 | 114 | 318 | 104 |
| 449 | 414 | 178 | 192 | 413 | 306 | 160 | 102 |
| 464 | 210 | 123 | 124 | 541 | 180 | 77 | 79 |
| 490 | 378 | 368 | 307 | 209 | 258 | 164 | 117 |
| 457 | 171 | 420 | 105 | 200 | 131 | 180 | 57 |
| 515 | 359 | 285 | 123 | 119 | 122 | 175 | 100 |
| 629 | 273 | 157 | 152 | 142 | 54 | 80 | 72 |



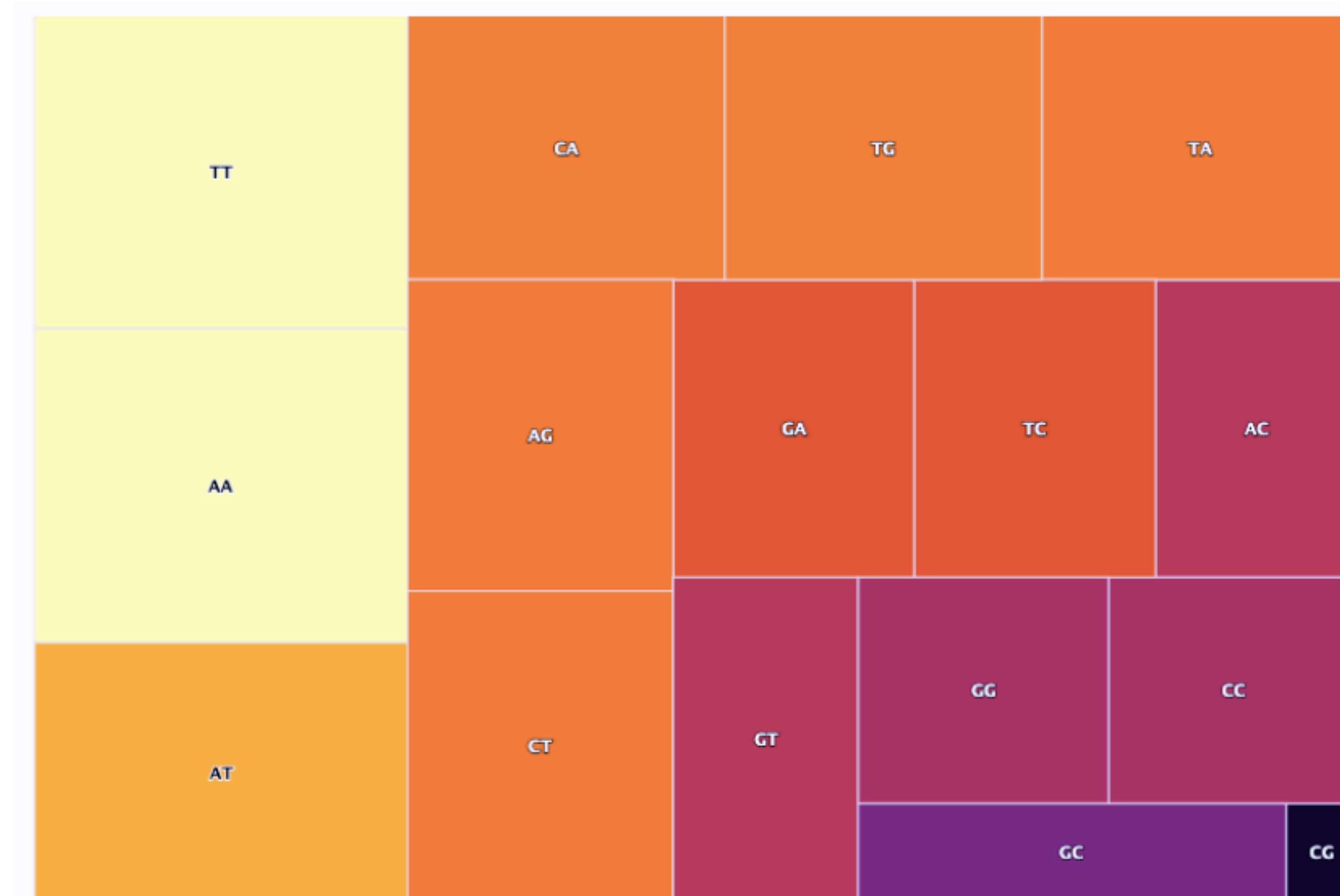
EN BACTERIAS CG ESTÁ ENRIQUECIDO COMPARADO CON EUCARIONTES

Escherichia coli cepa K-12

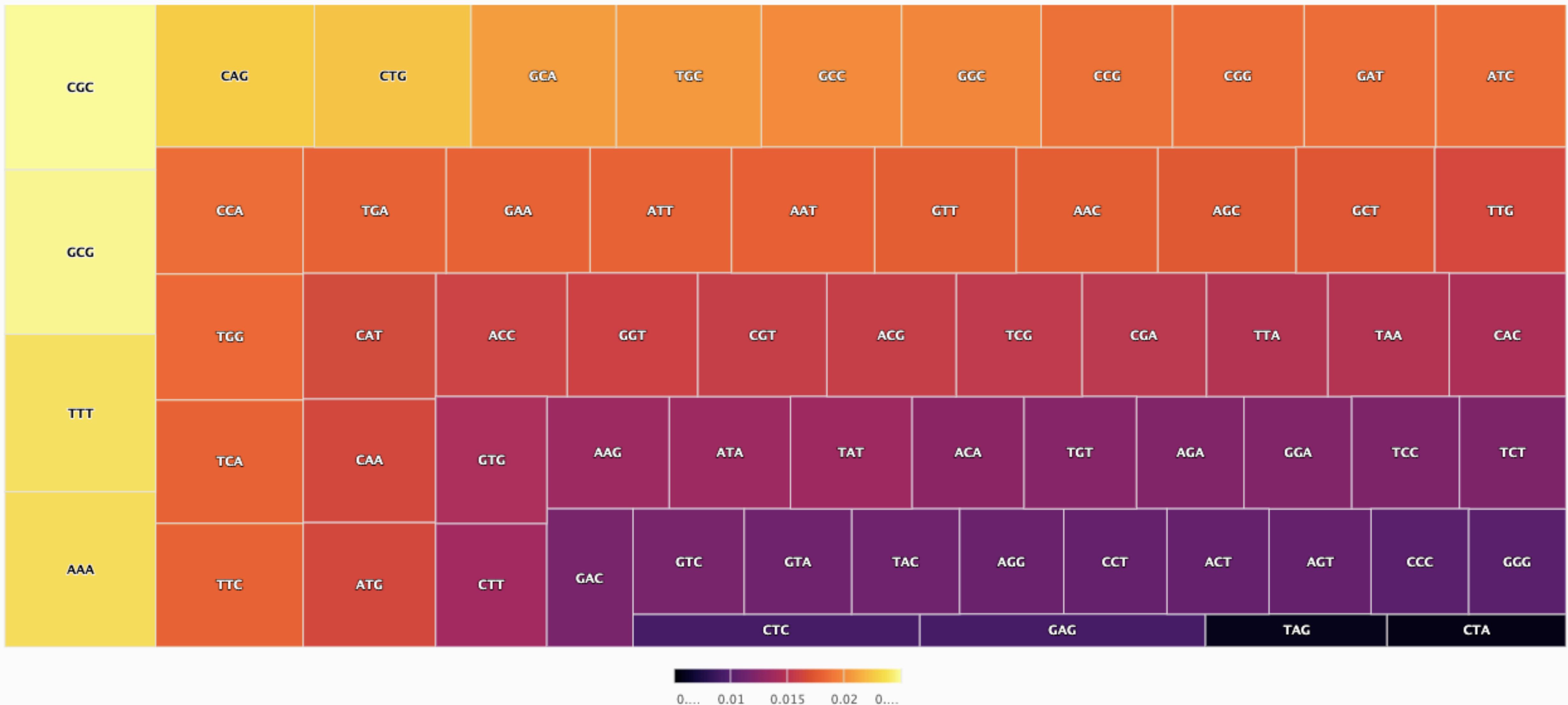


EN BACTERIAS CG ESTÁ ENRIQUECIDO COMPARADO CON EUCARIONTES

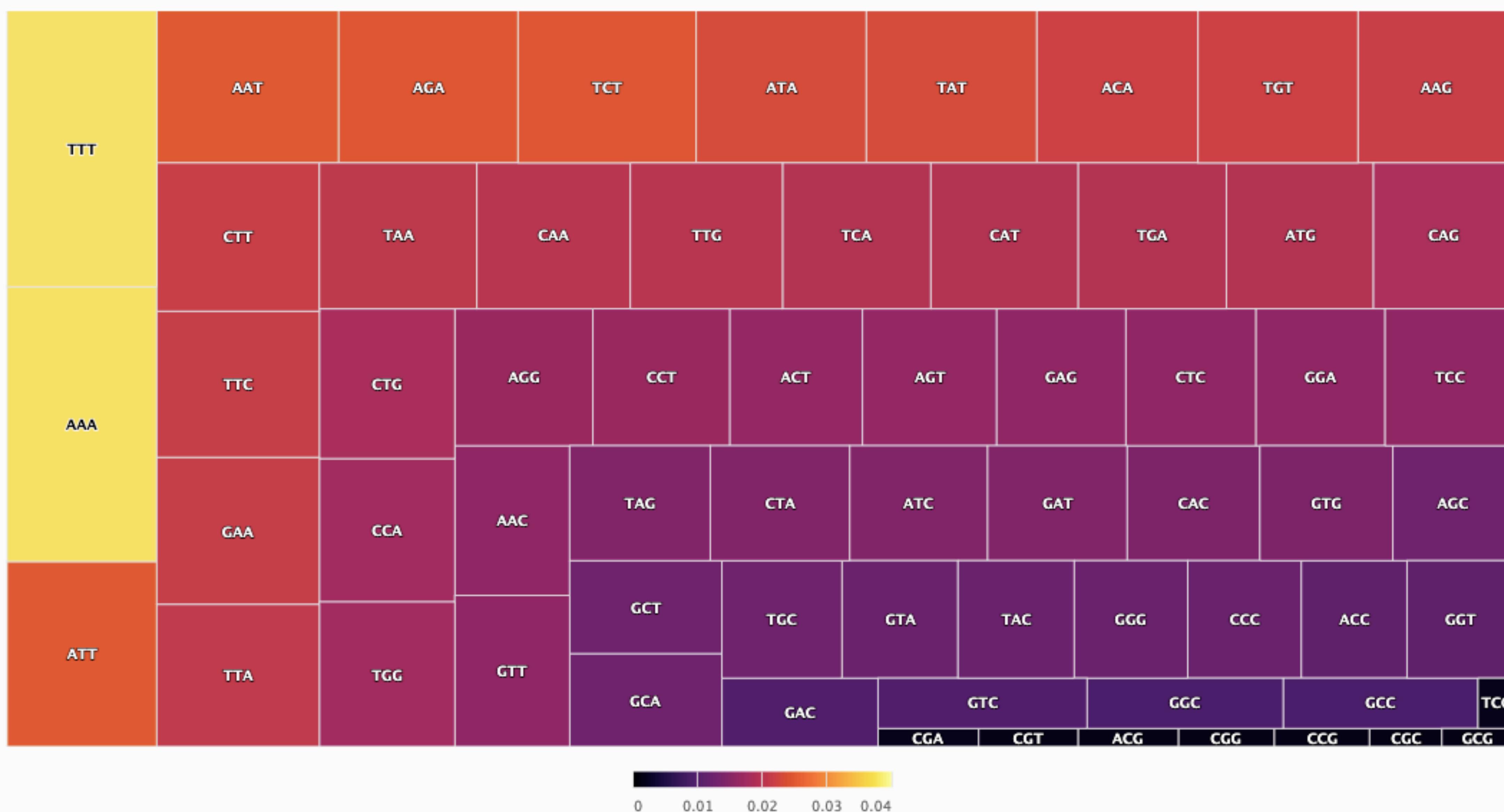
Ratón Cromosoma X



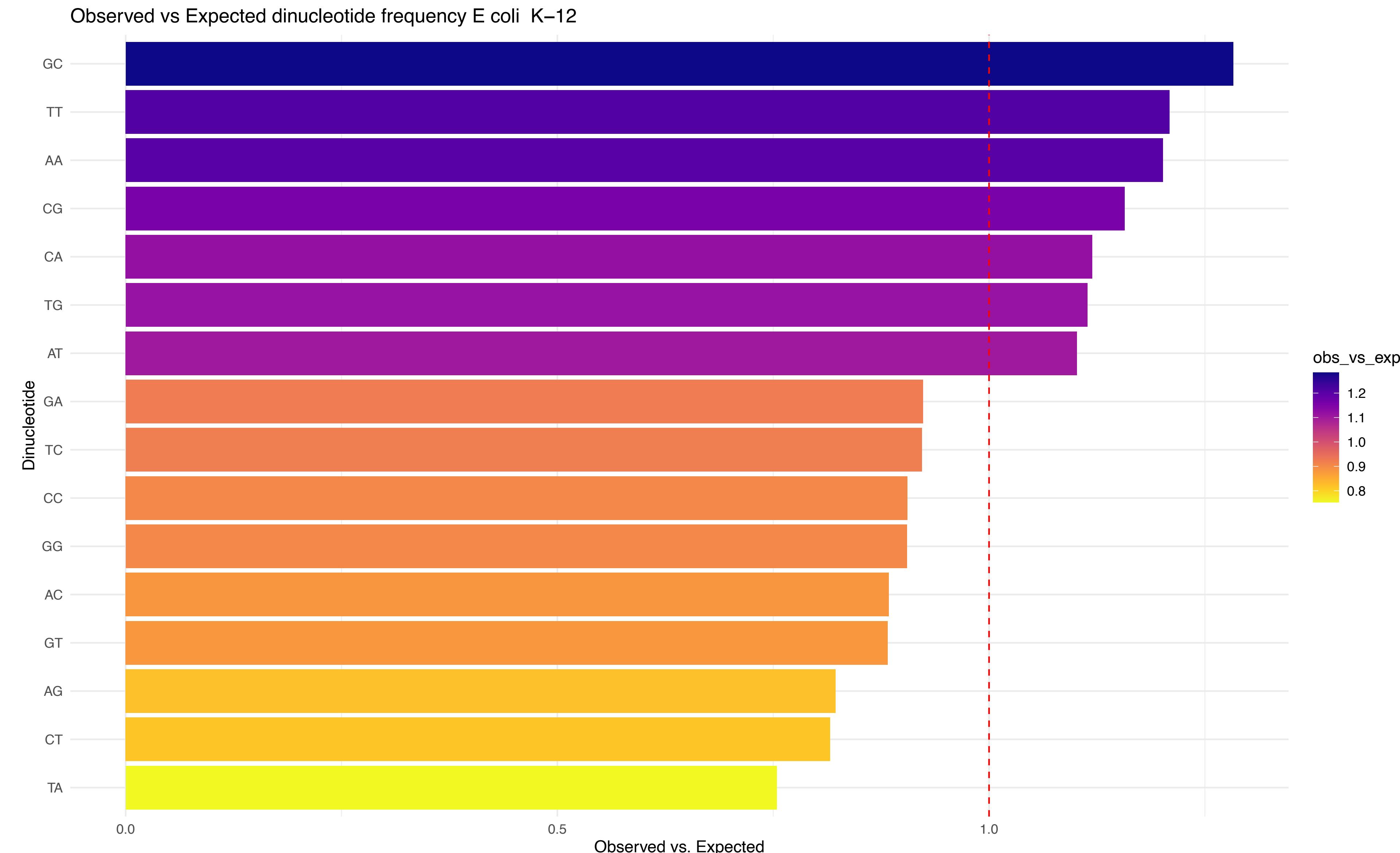
EN BACTERIAS CG ESTÁ ENRIQUECIDO COMPARADO CON EUCARIONTES



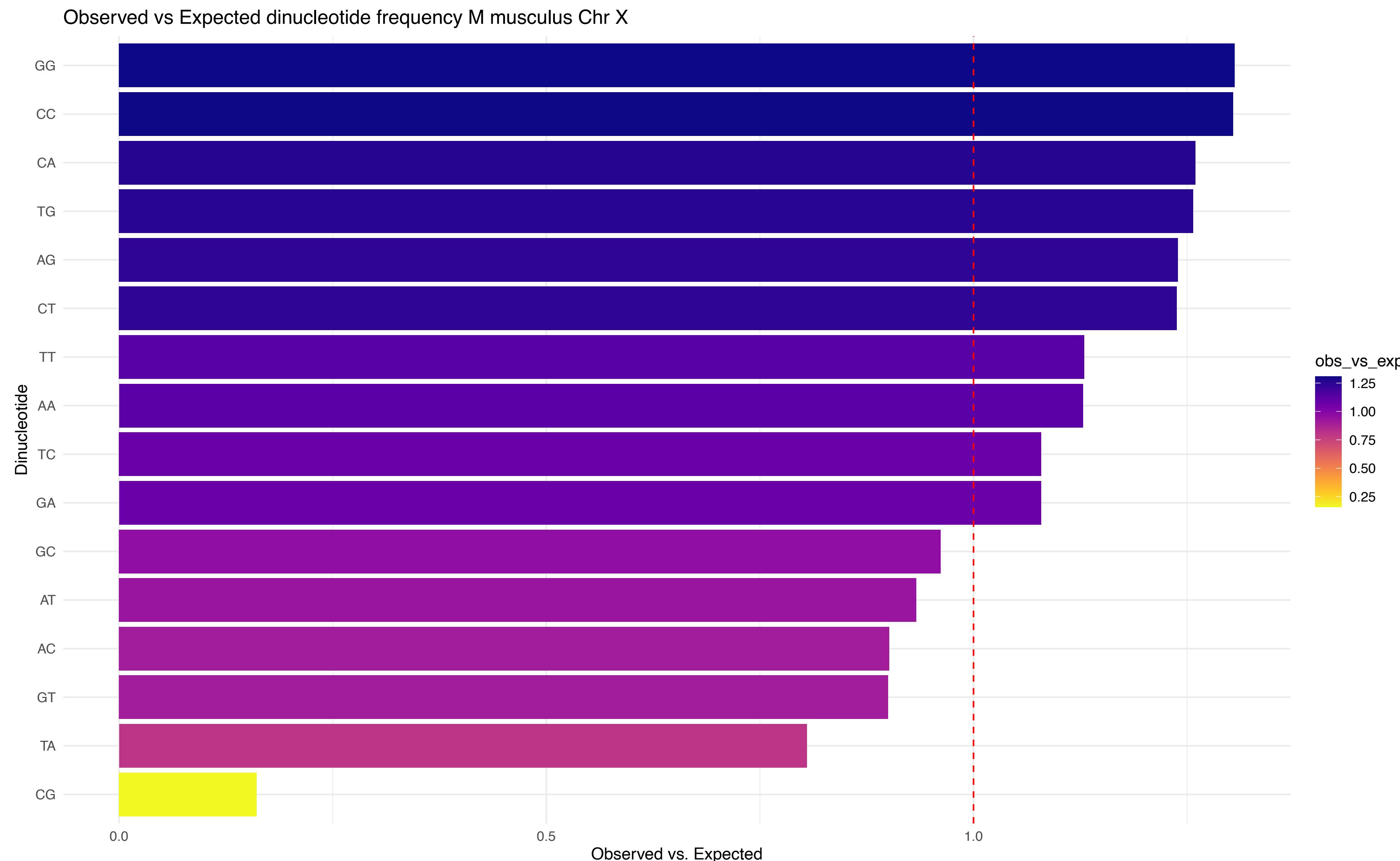
EN BACTERIAS CG ESTÁ ENRIQUECIDO COMPARADO CON EUCARIONTES



EN BACTERIAS CG APARECE 30 % MÁS DE LO ESPERADO



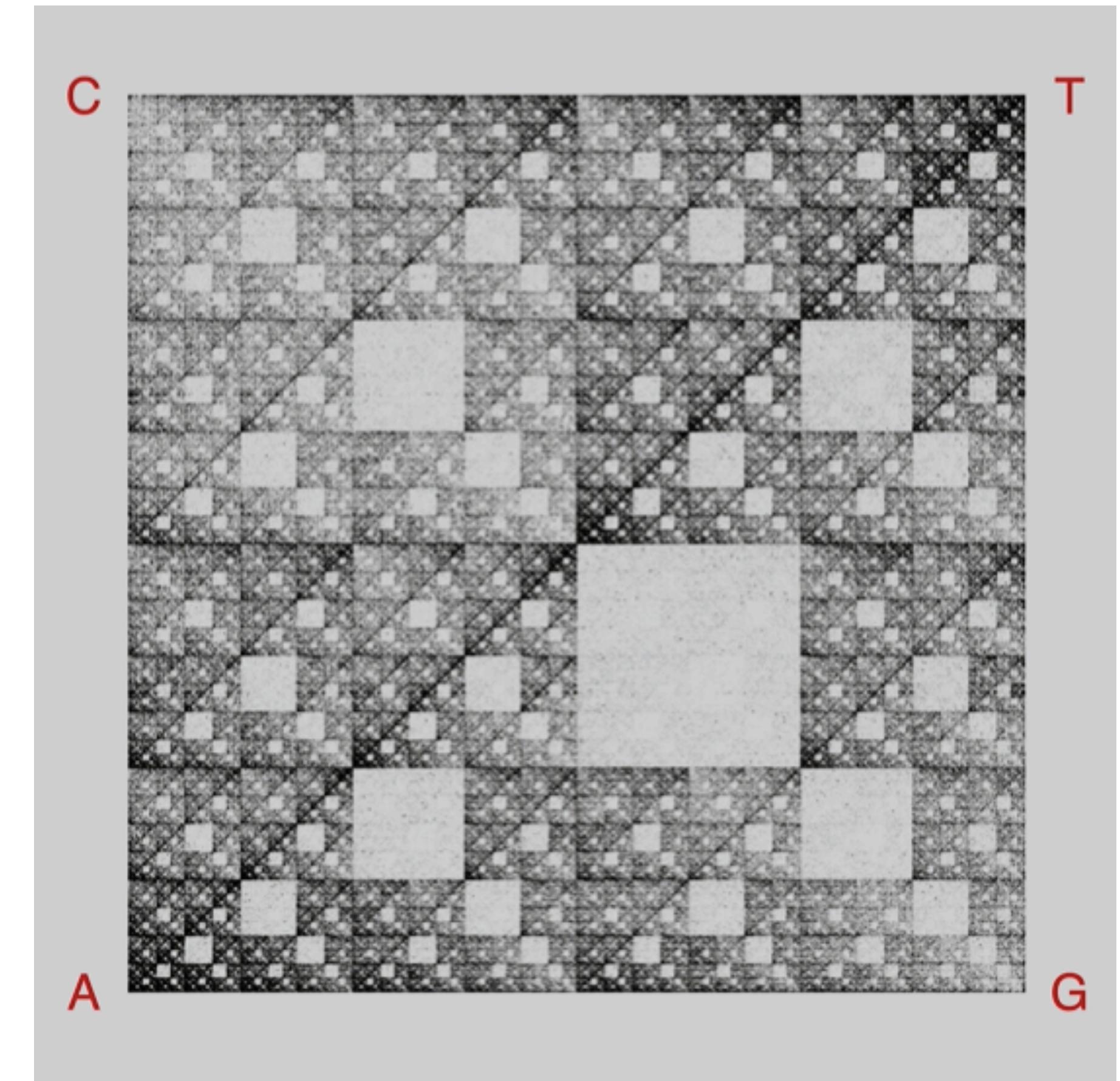
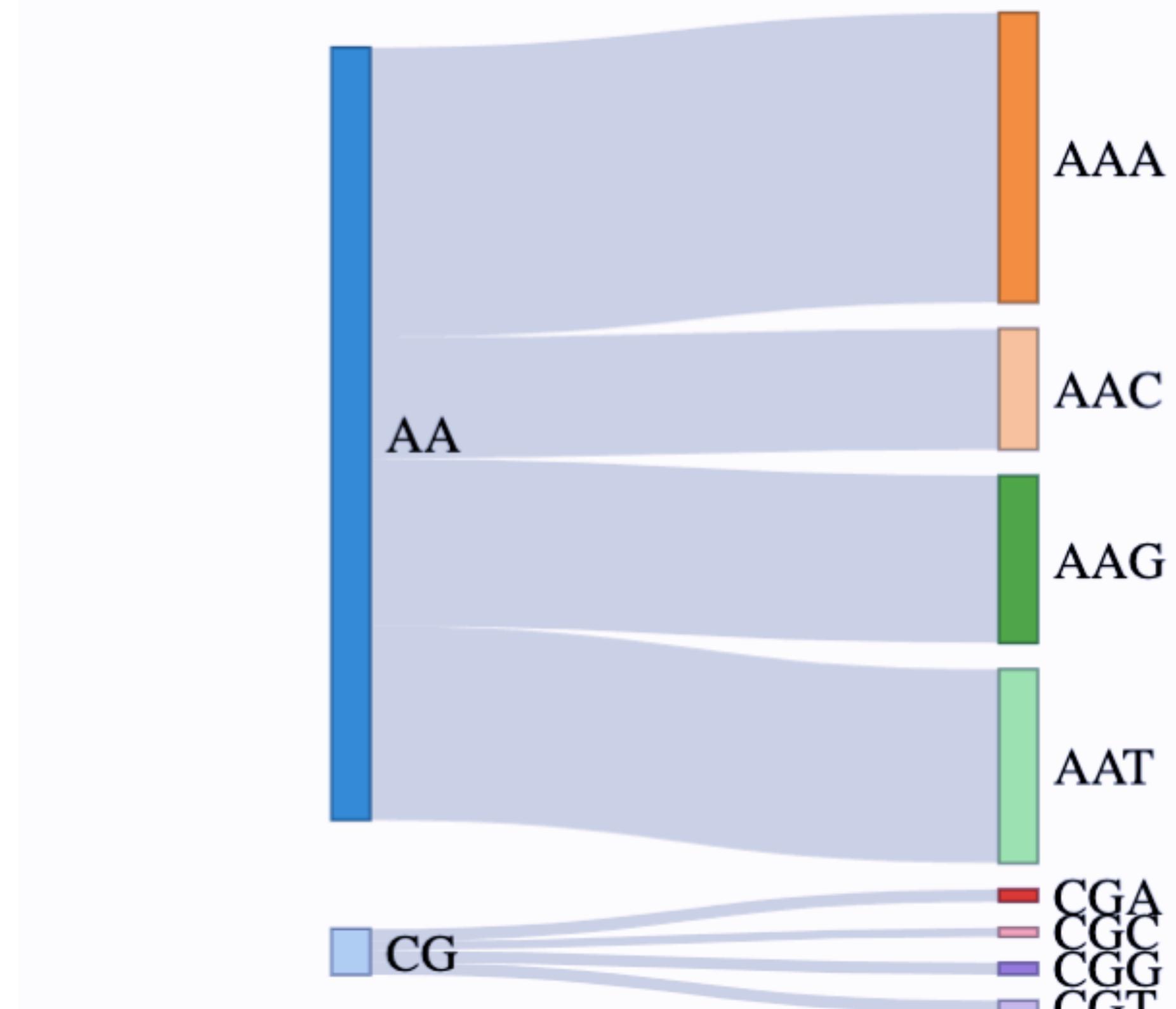
CG SE OBSERVA MENOS DE 5 VECES DE LO ESPERADO



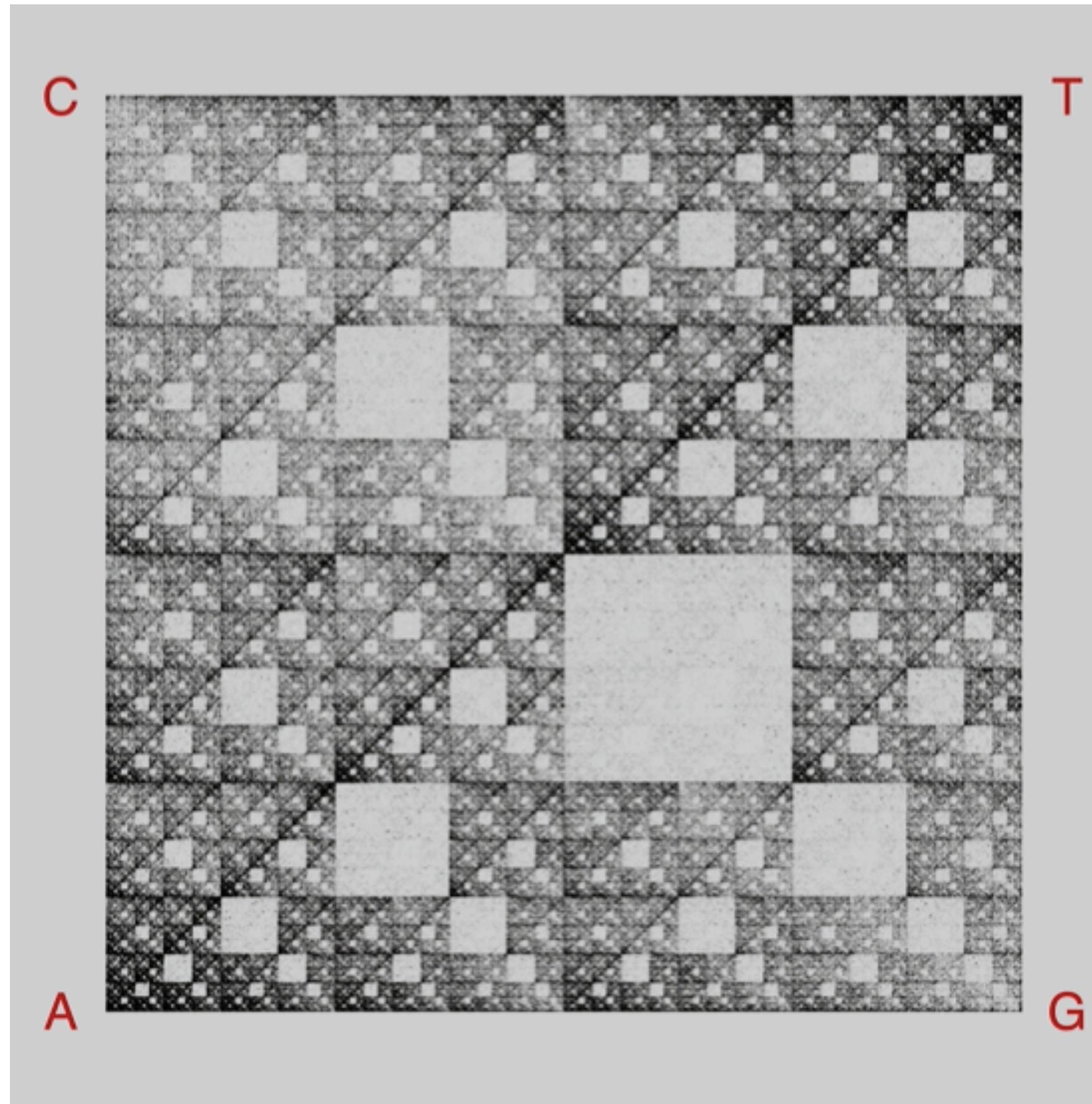
LOS FRACTALES SE OBTIENEN POR LA FRECUENCIA DE N-OMEROS

LOS PATRONES SE HEREDAN POR N-OMEROS DE ORDEN MÁS BAJO

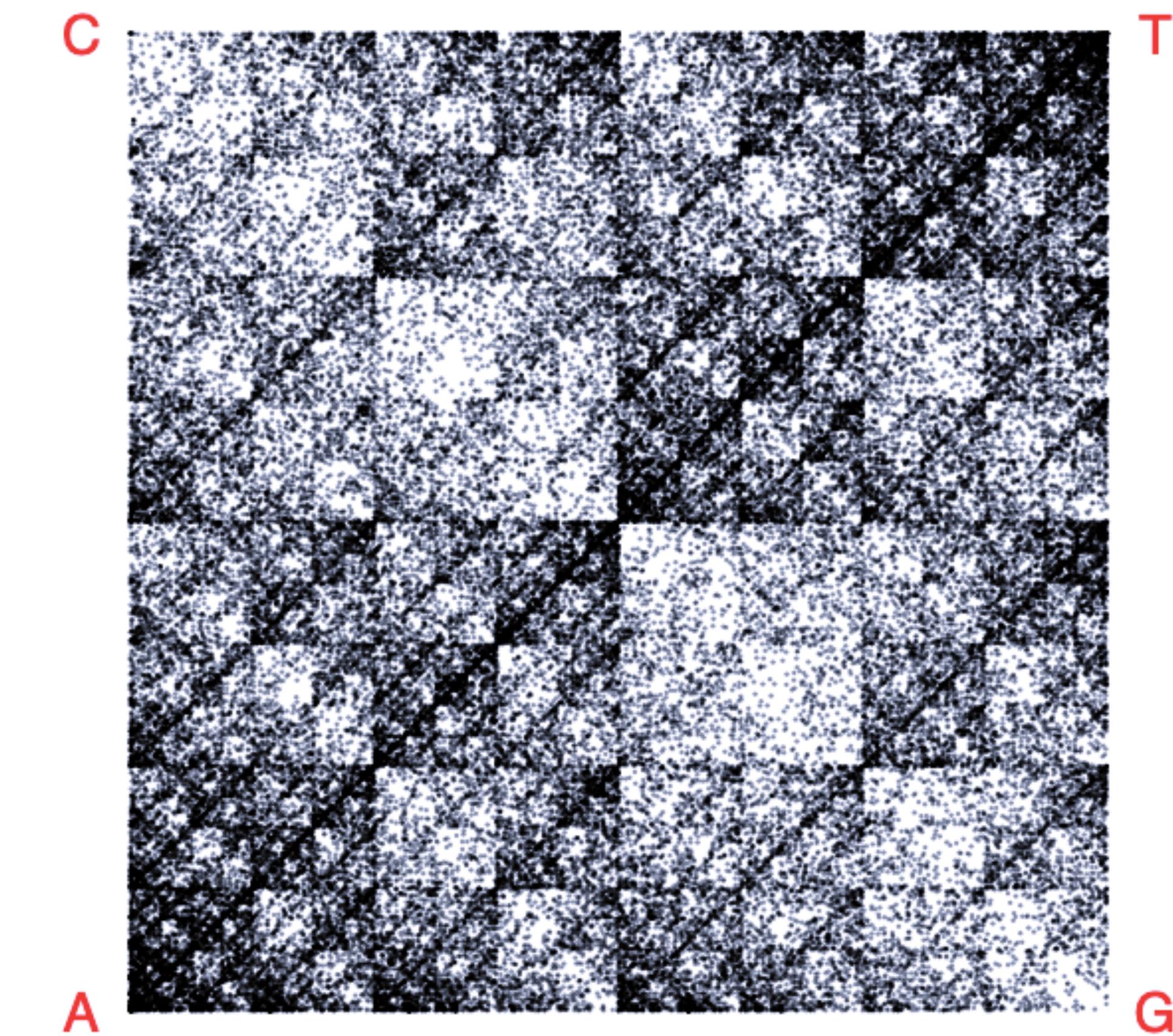
Ratón



FRACTALES EN EL DNA



Mus musculus

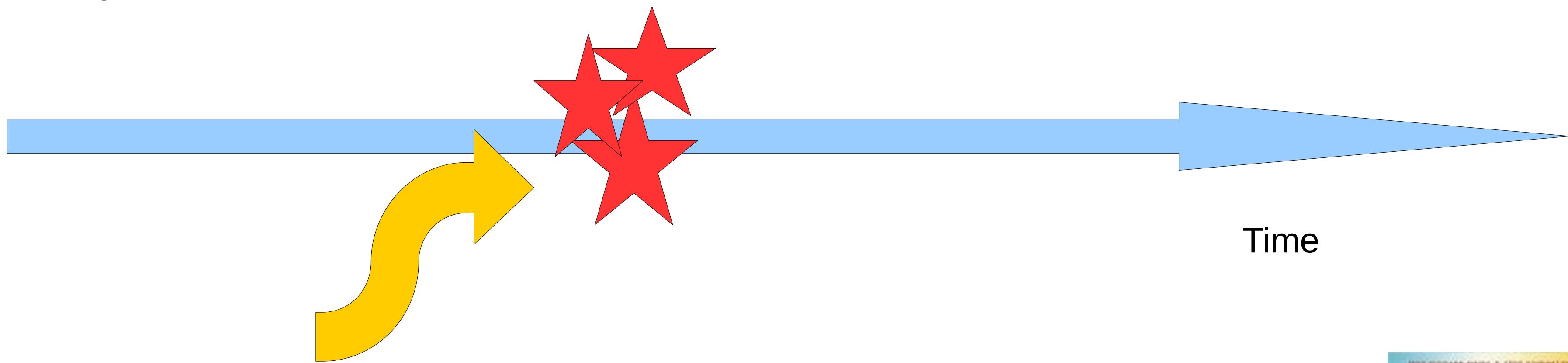


Arabidopsis thaliana

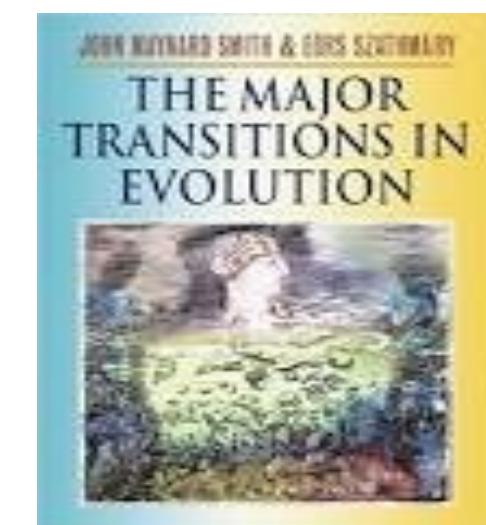
DISTINTAS FRECUENCIAS DE APARICIÓN

In Eukaryotes CpG is underrepresented

In Prokaryotes CpG is usually overrepresented



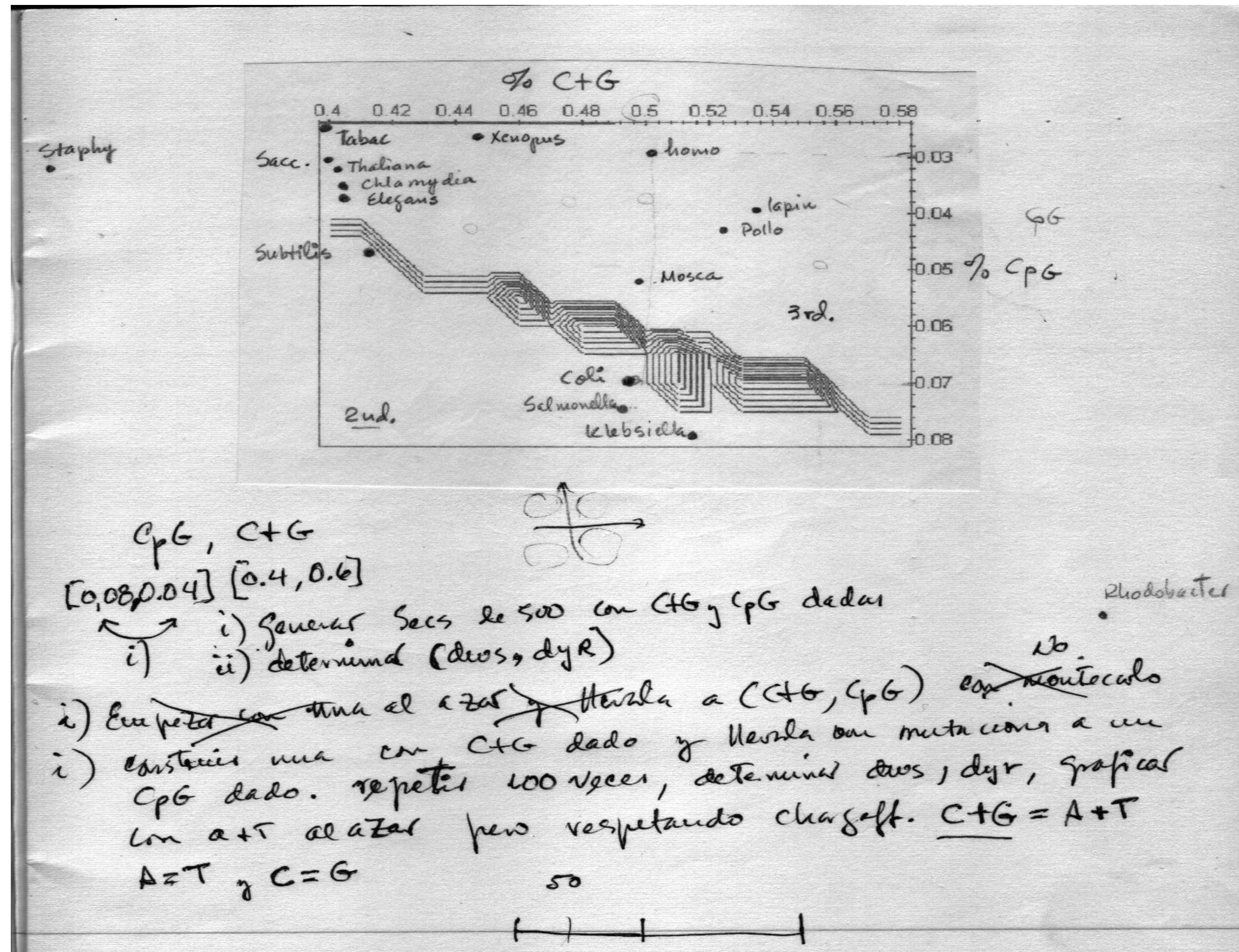
A major transition in evolution

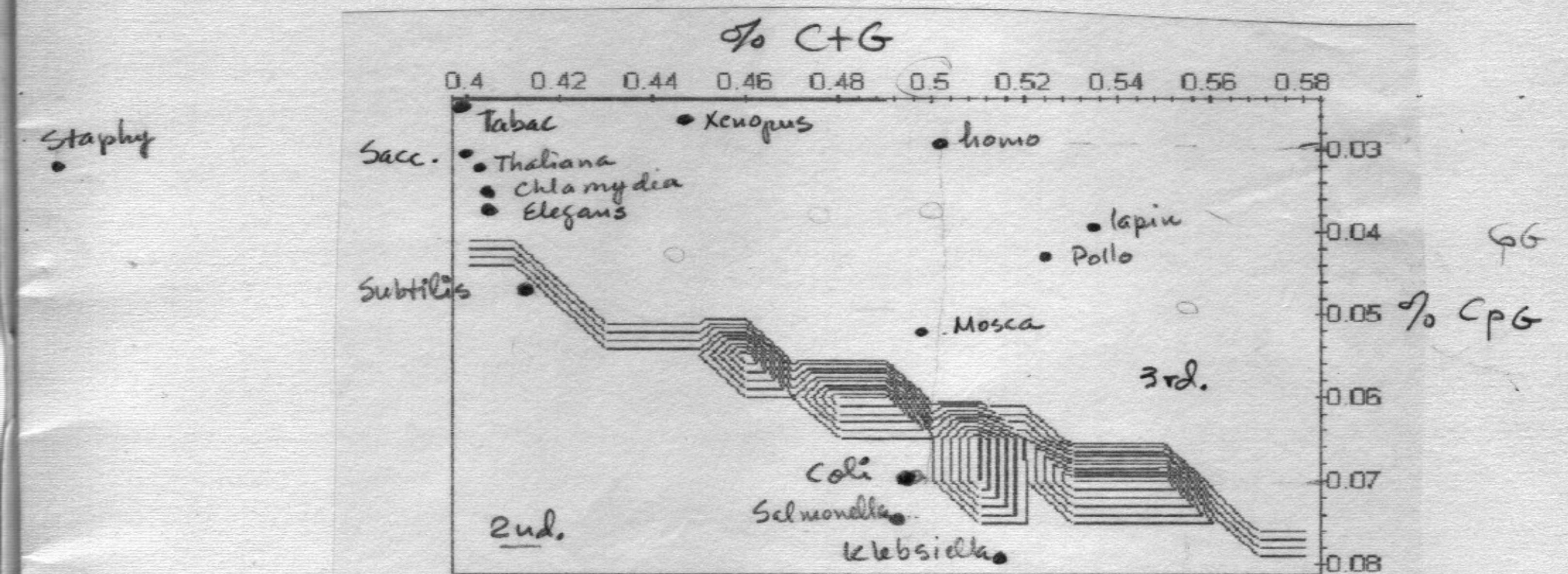


IDEA CENTRAL

The Emergence of Eukaryotes

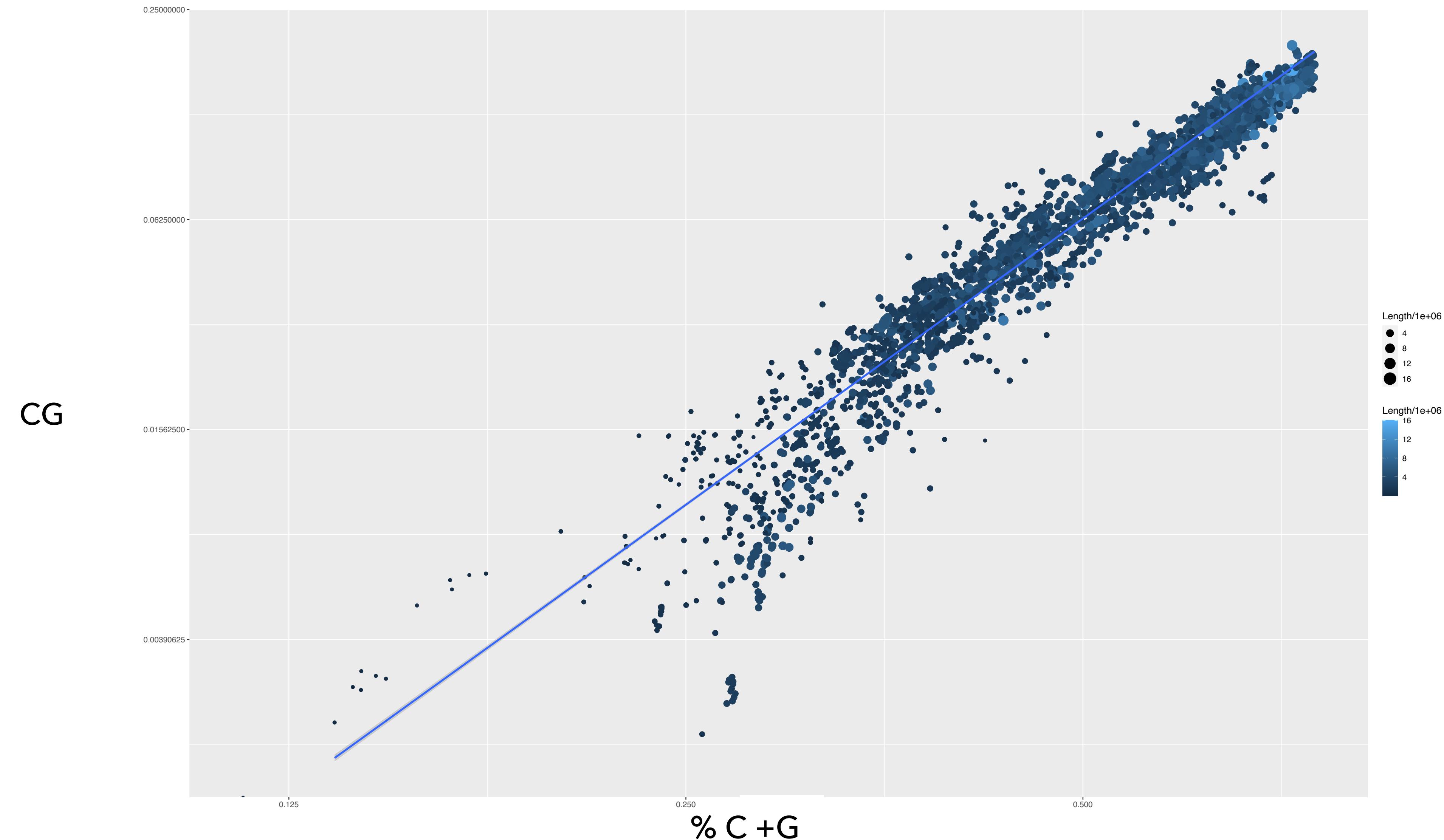
A phase transition??





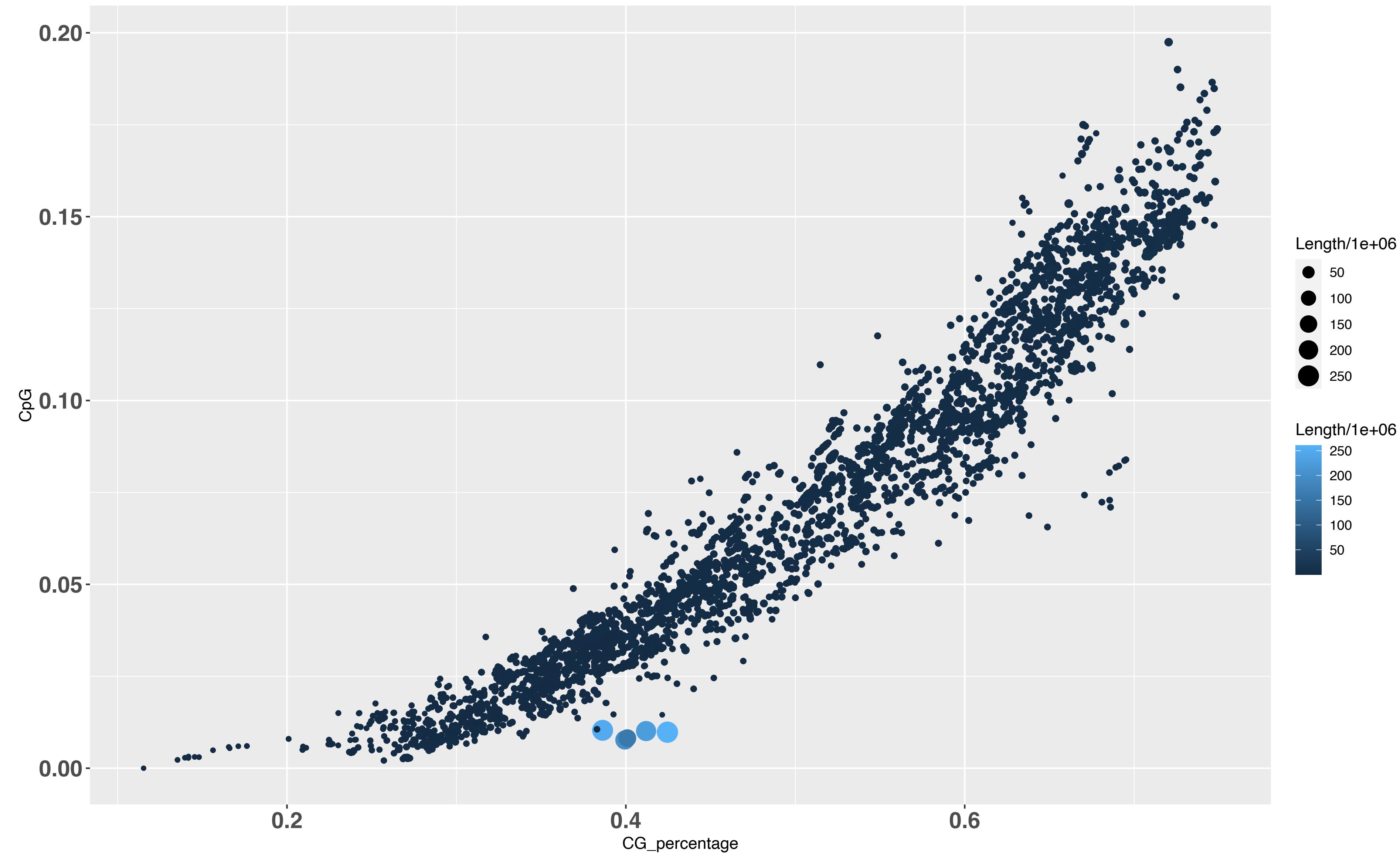
- $CpG, C+G$
 $[0.08, 0.04] [0.4, 0.6]$
- i) Generar Secs de 500 con $C+G$ y CpG dados
 ii) determinar α os, δ gr, graficar
- i) ~~Especie~~ con otra el ~~azar~~ → lleva a $(C+G, CpG)$ ~~con~~ montecarlo
 ii) construir una con $C+G$ dado y lleva con mutaciones a un CpG dado. repetir 100 veces, determinar α os, δ gr, graficar
 con $A=T$ al azar pero respetando chargeff. $\underline{C+G = A+T}$
 $A=T$ y $C=G$

MILES DE BACTERIAS EN UN AJUSTE EXPONENCIAL



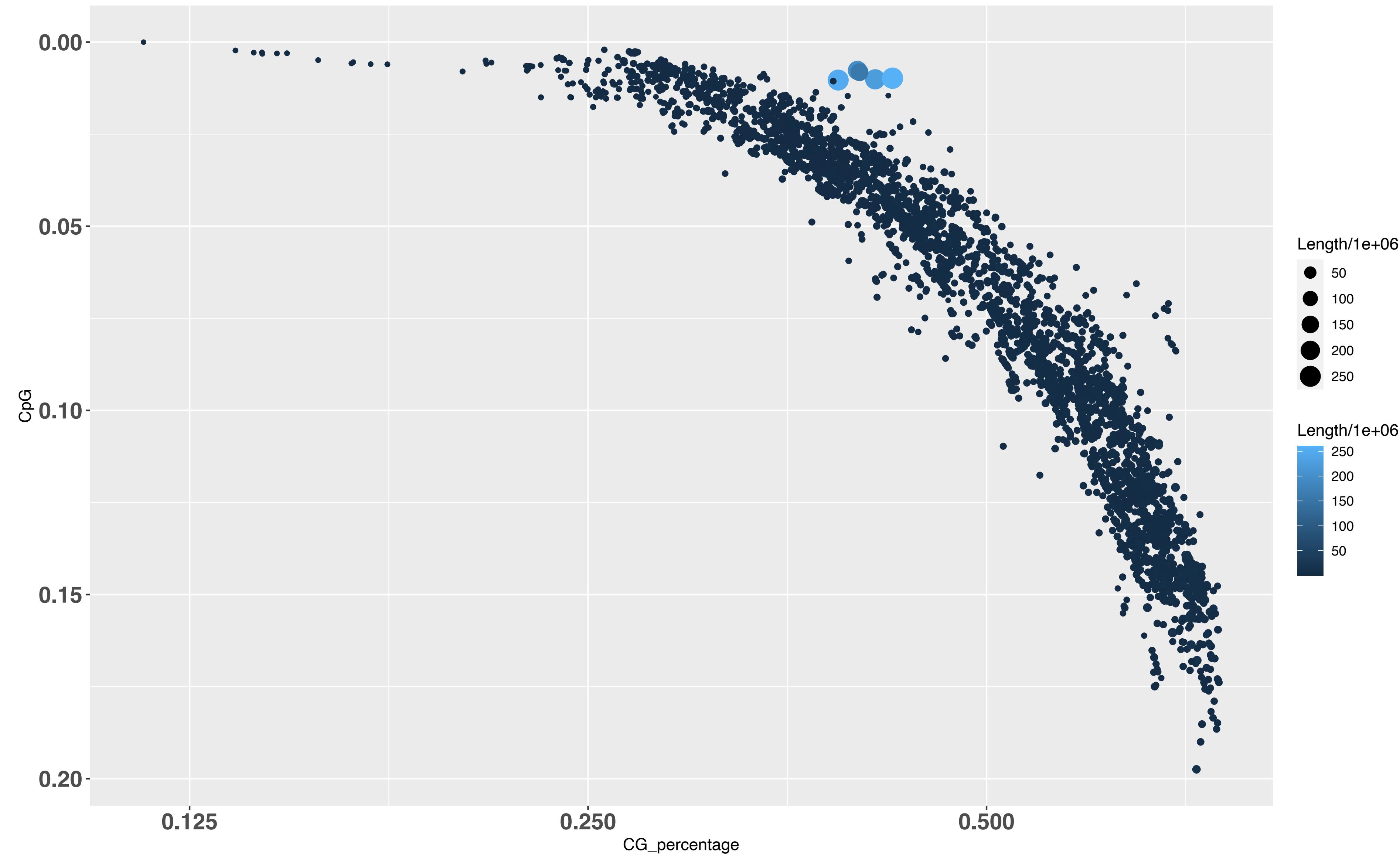
ZONAS

DIAGRAMA DE FASE



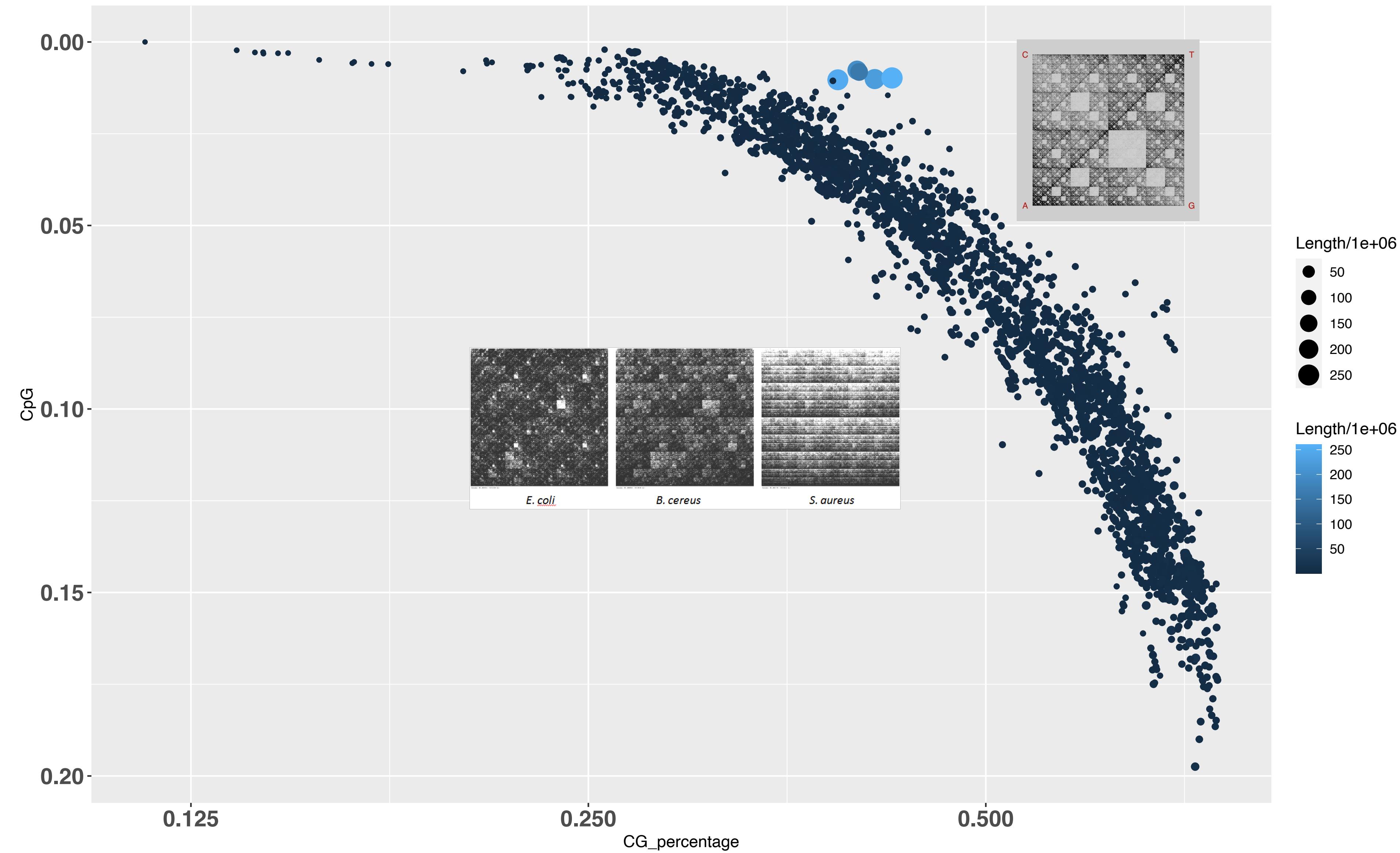
ZONAS

DIAGRAMA DE FASES



ZONAS

DIAGRAMA DE FASES



HIPÓTESIS

- ▶ Las diferentes frecuencias de aparición de los n-omeros se debe a un “potencial de interacción” que favorece a ciertos “estados” sobre otros.
- ▶ Estados más (menos) frecuentes (probables) tienen menos (más) “energía”

FRECUENCIAS

$$p_i(s_1 s_2 \dots s_n)$$

$$s_j \in \{A, C, G, T\}$$

$$i = 1, 2, \dots, 4^n$$

FRECUENCIAS

$$p_i(s_1 s_2 \dots s_n)$$

$$H(s_1 s_2 \dots s_n) = -\ln(p_i)$$

$$\epsilon_i \equiv H(p_i)$$

DISTRIBUCIÓN ESCORT

$$H(s_1 s_2 \dots s_n) = -\log(p_i)$$

$$\epsilon_i \equiv H(p_i)$$

$$\mathcal{P}_i = \frac{p_i^\nu}{\sum_i^{4^n} p_i}$$

DISTRIBUCIÓN ESCORT

$$\epsilon_i \equiv H(p_i) = -\log(p_i)$$

$$\mathcal{P}_i = \frac{p_i^\nu}{\sum_i^{4^n} p_i}$$

$$e^{-\epsilon_i \nu} = p_i^\nu$$

$$\mathcal{P}_i = \frac{e^{-\epsilon_i \nu}}{\sum_{i=1}^{4^n} e^{-\epsilon_i \nu}}$$

$$\mathcal{Z} = \sum_{i=1}^{4^n} e^{-\epsilon_i \nu}$$

$$\nu \rightarrow \frac{1}{KT}$$

ENTROPÍA GENERALIZADA

$$K_\nu = \frac{1}{1-\nu} \ln \sum p^\nu$$

$$\nu = 0 \ K_0 = \log(4^n)$$

$$\nu = 1 \ K_1 = - \sum_i p_i \ln(p_i)$$

ENERGÍA LIBRE

$$\tau(\nu) = -\frac{1}{\nu} \ln \mathcal{Z}$$

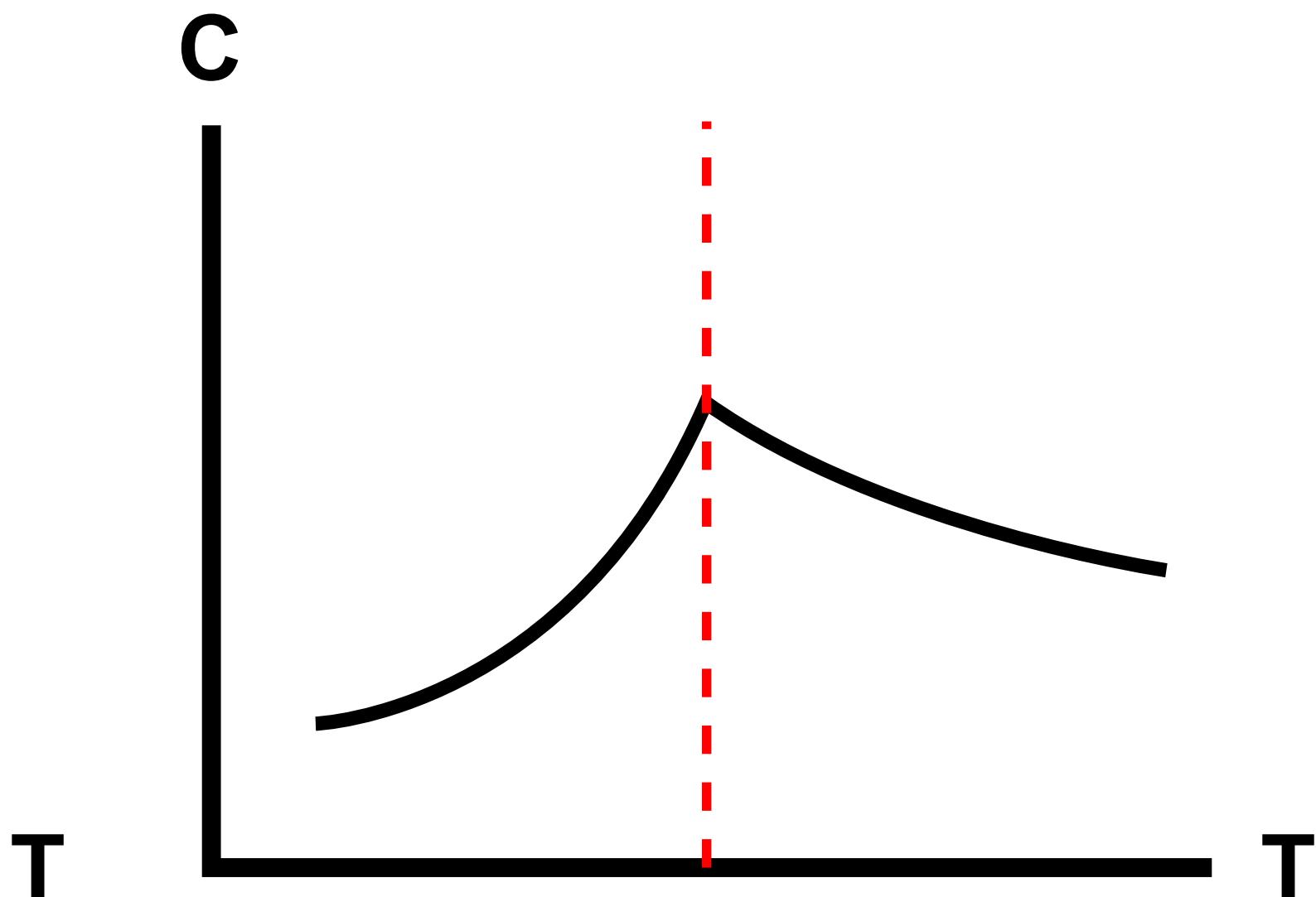
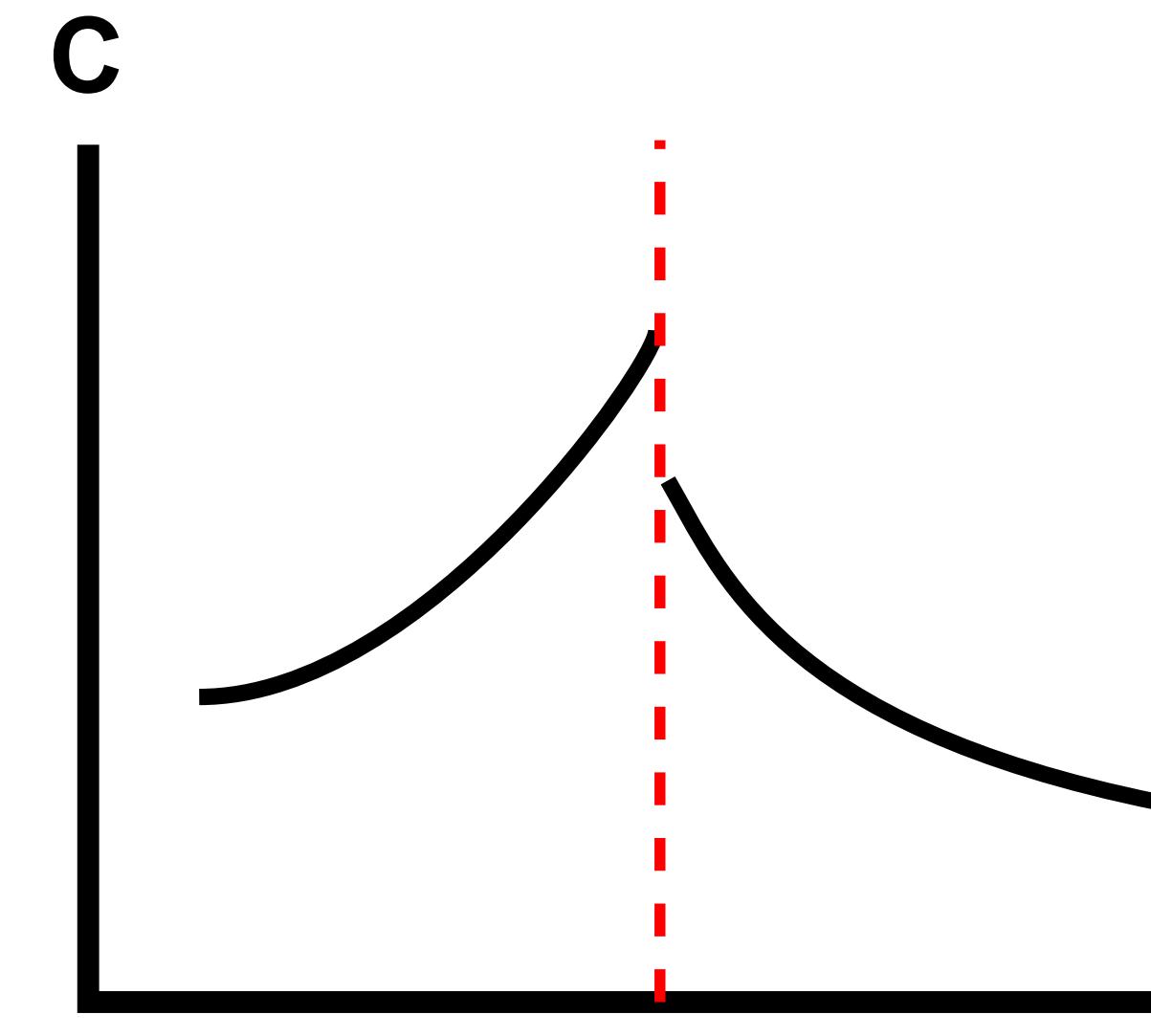
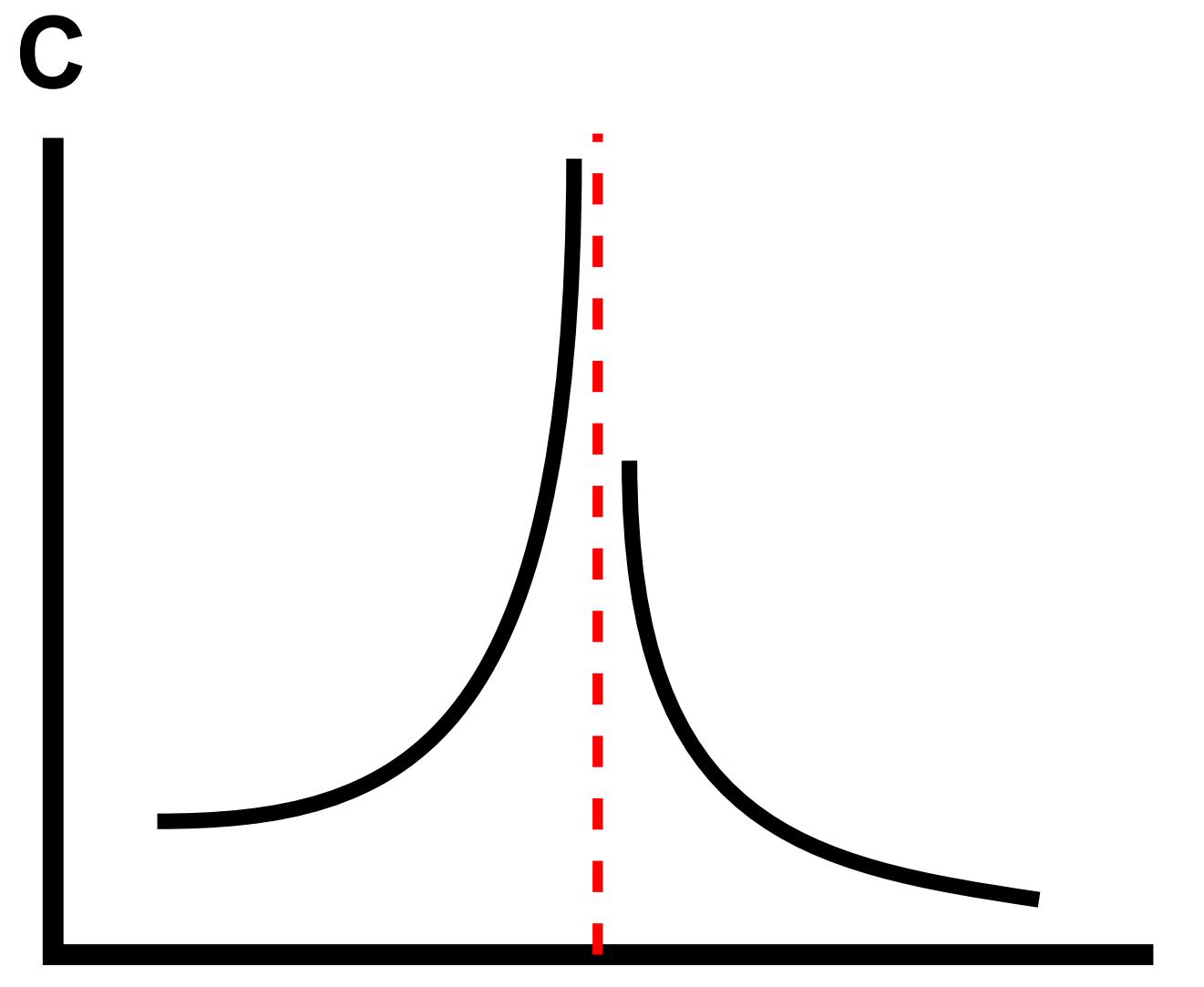
$$\tau(\nu) = (\nu - 1)K_\nu$$

$$C(\nu) = \frac{\partial^2 \tau(\nu)}{\partial \nu^2} \approx 2\tau(\nu) - \tau(\nu + 1) - \tau(\nu - 1)$$

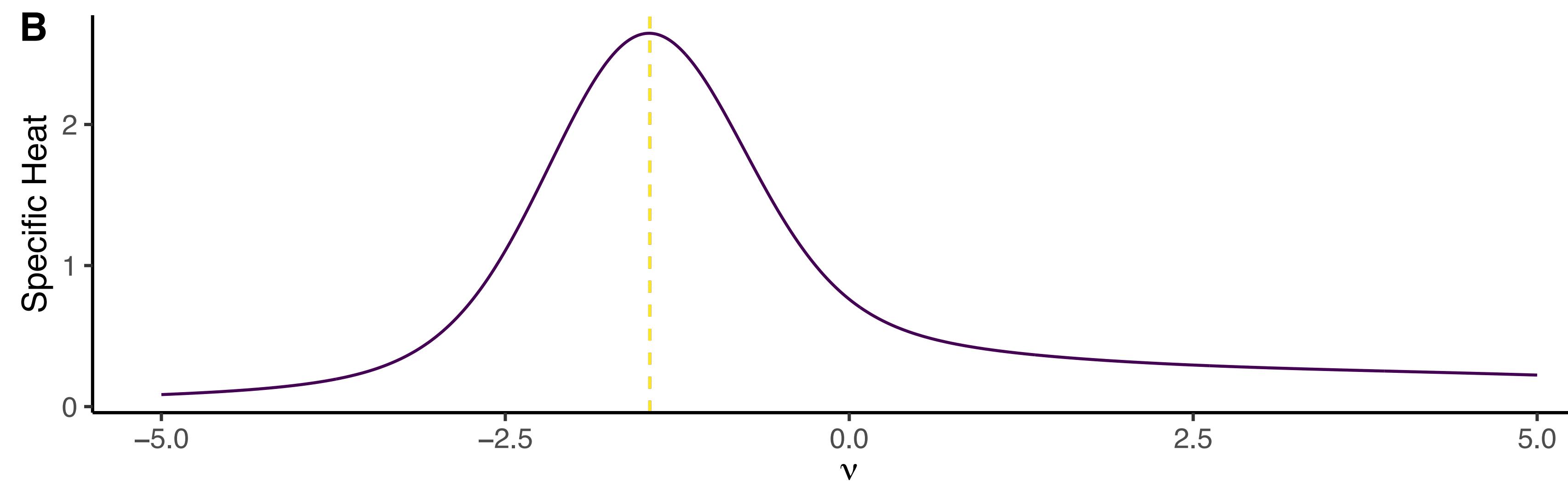
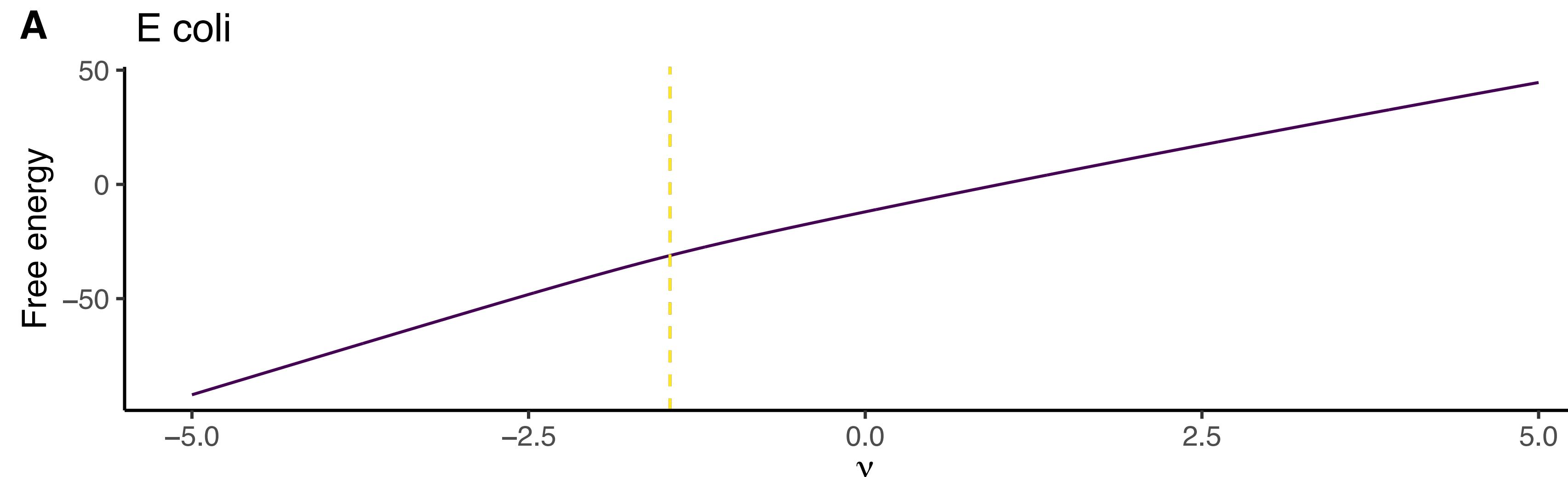
$$C(\nu) \approx 2(\nu - 1)K_\nu - \nu K_{\nu+1} - (\nu - 1)K_{\nu-1}$$

¿CÓMO IDENTIFICAR TRANSICIONES DE FASE?

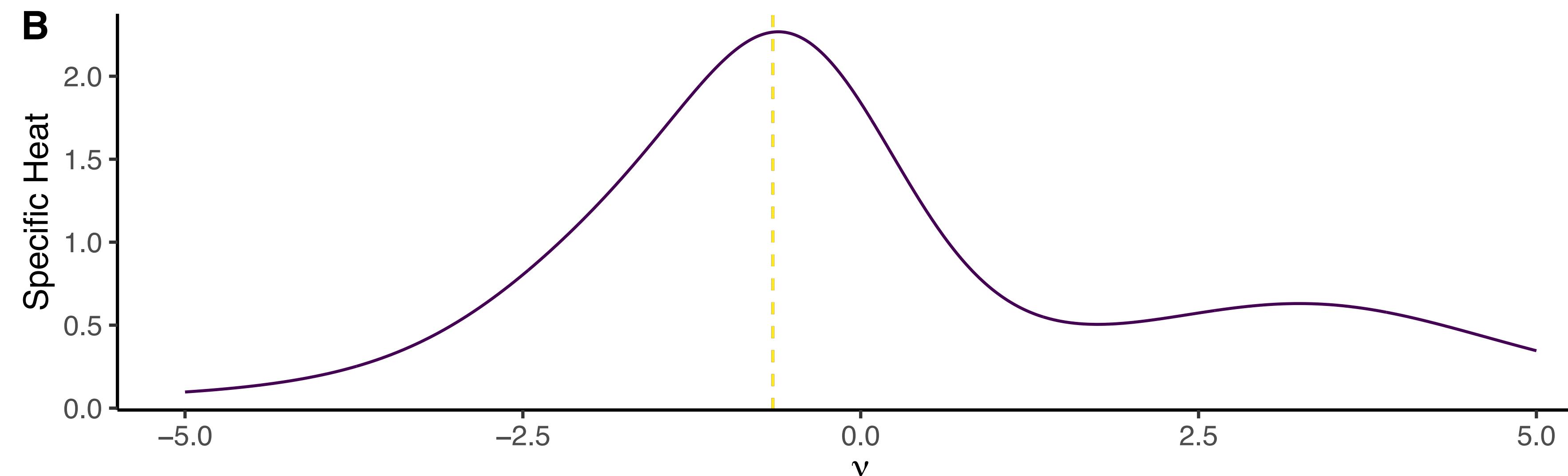
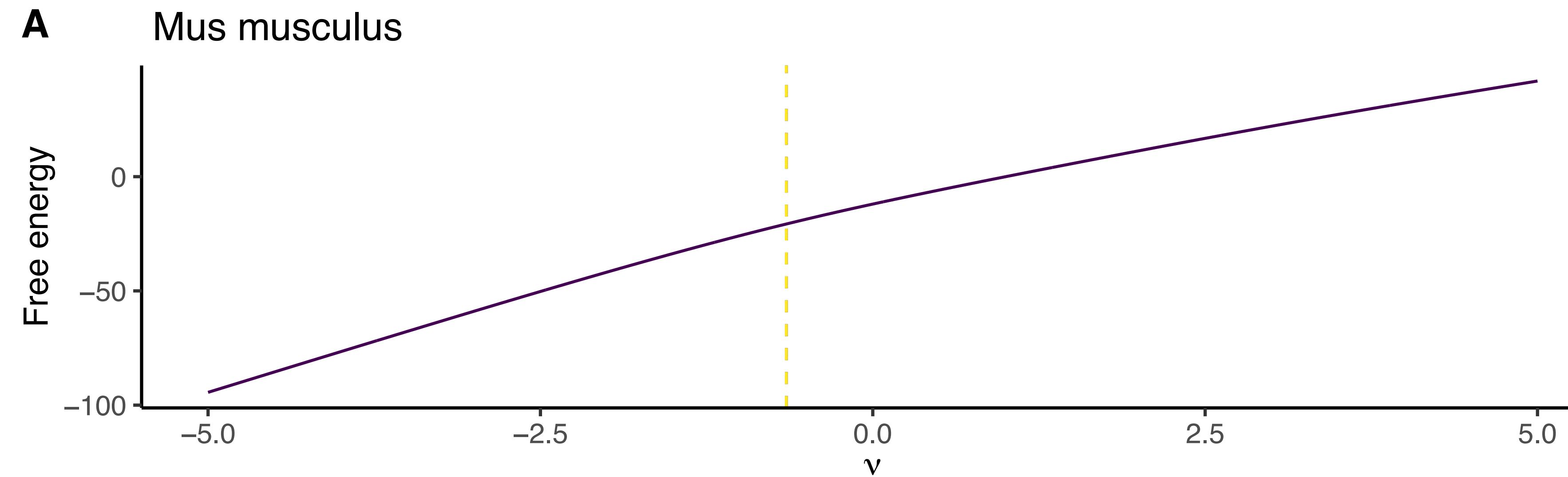
LAS TRANSICIONES DE FASE SE OBSERVAN COMO DISCONTINUIDADES DE C



FRECUENCIAS EMPÍRICAS

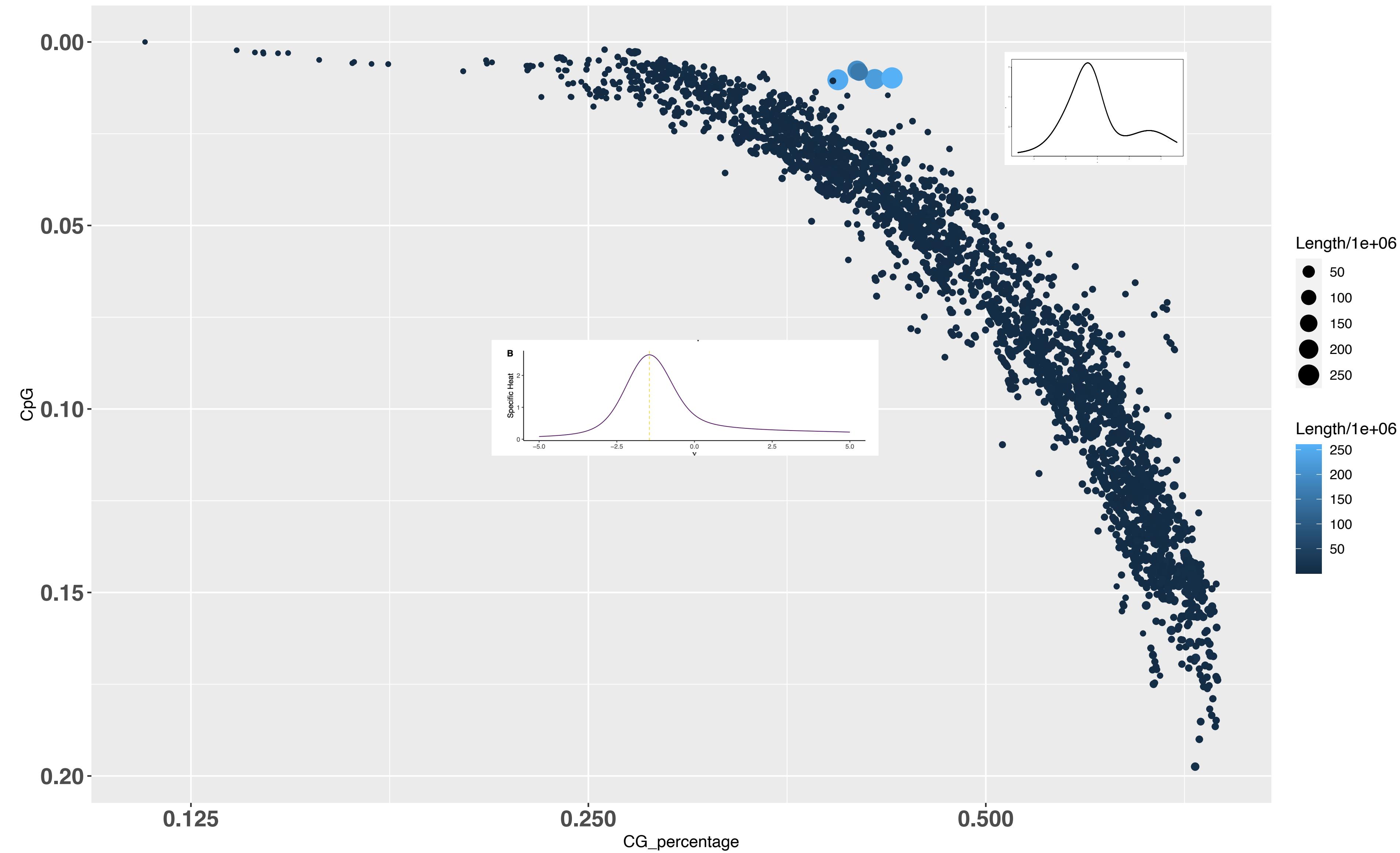


FRECUENCIAS EMPÍRICAS

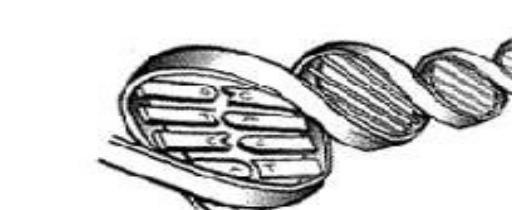
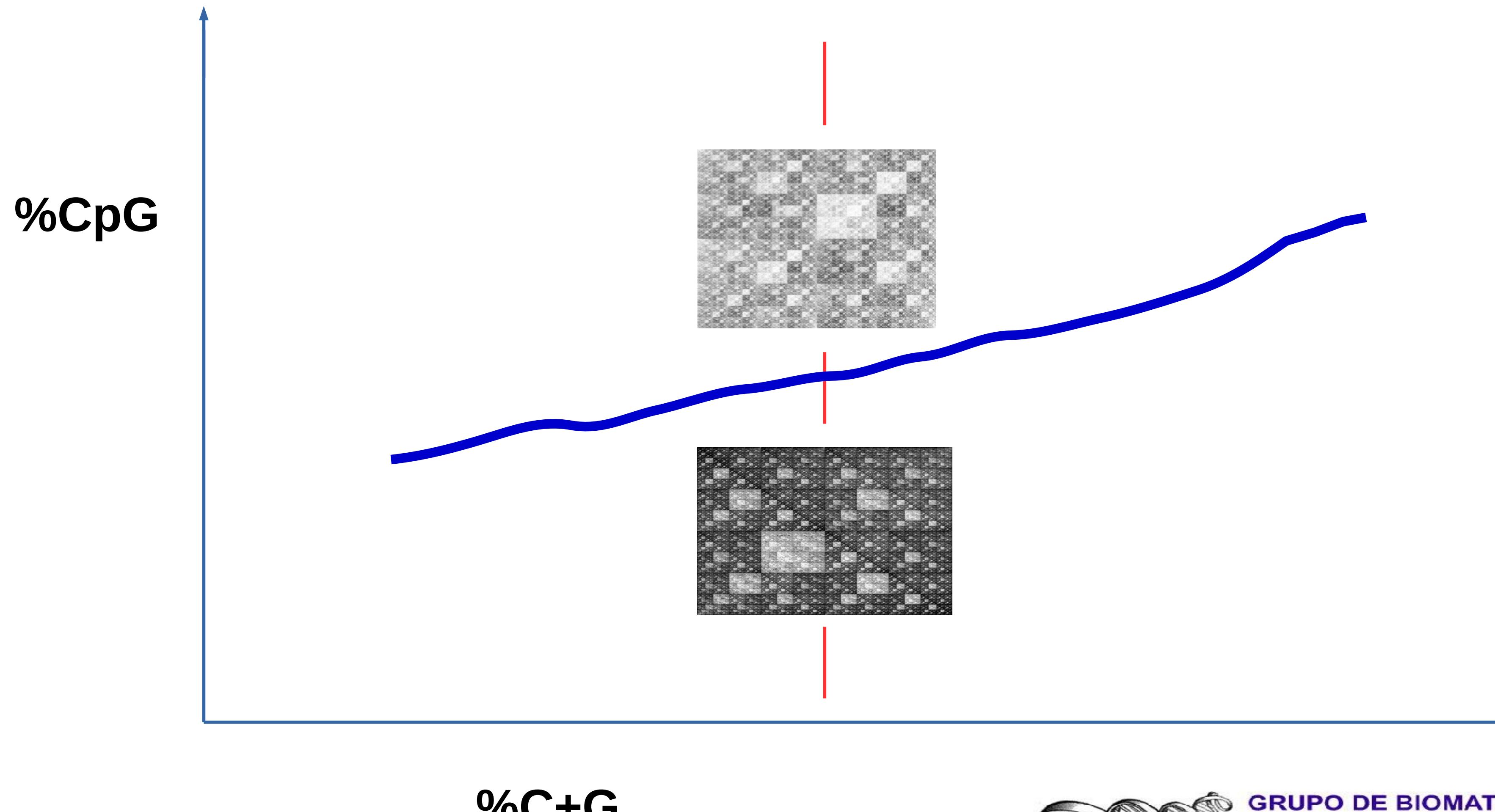


ZONAS

DIAGRAMA DE FASES



¿COMO PASAR DE UN PATRÓN FRACTAL A OTRO?



MATRICES DE MARKOV

En los organismos multicelulares CpG está subrepresentado.

En las bacterias CpG suele estar sobrerepresentado

$$\begin{bmatrix} aa & ac & ag & at \\ ca & cc & cg & ct \\ ga & gc & gg & gt \\ ta & tc & tg & tt \end{bmatrix}$$

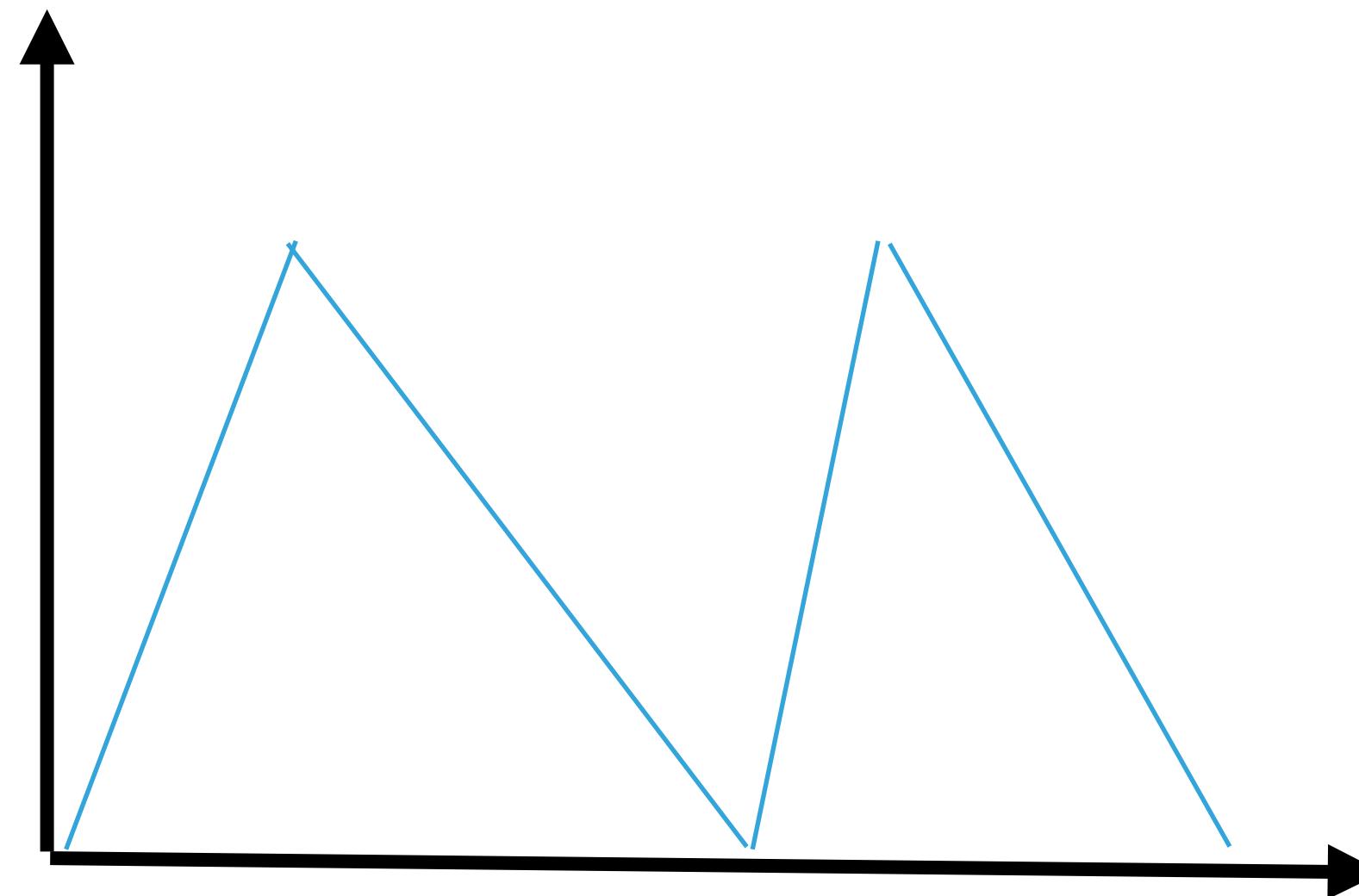
EN BACTERIAS SE PUEDE GENERAR UN MAPEO QUE NOS DA LA ENERGÍA LIBRE

$$K_q = \frac{1}{1-q} \ln(w_1^q + w_2^q + \dots w_k^q), \quad q \neq 1$$

$$K_1 = \sum_{i=1}^k w_i \ln w_i$$

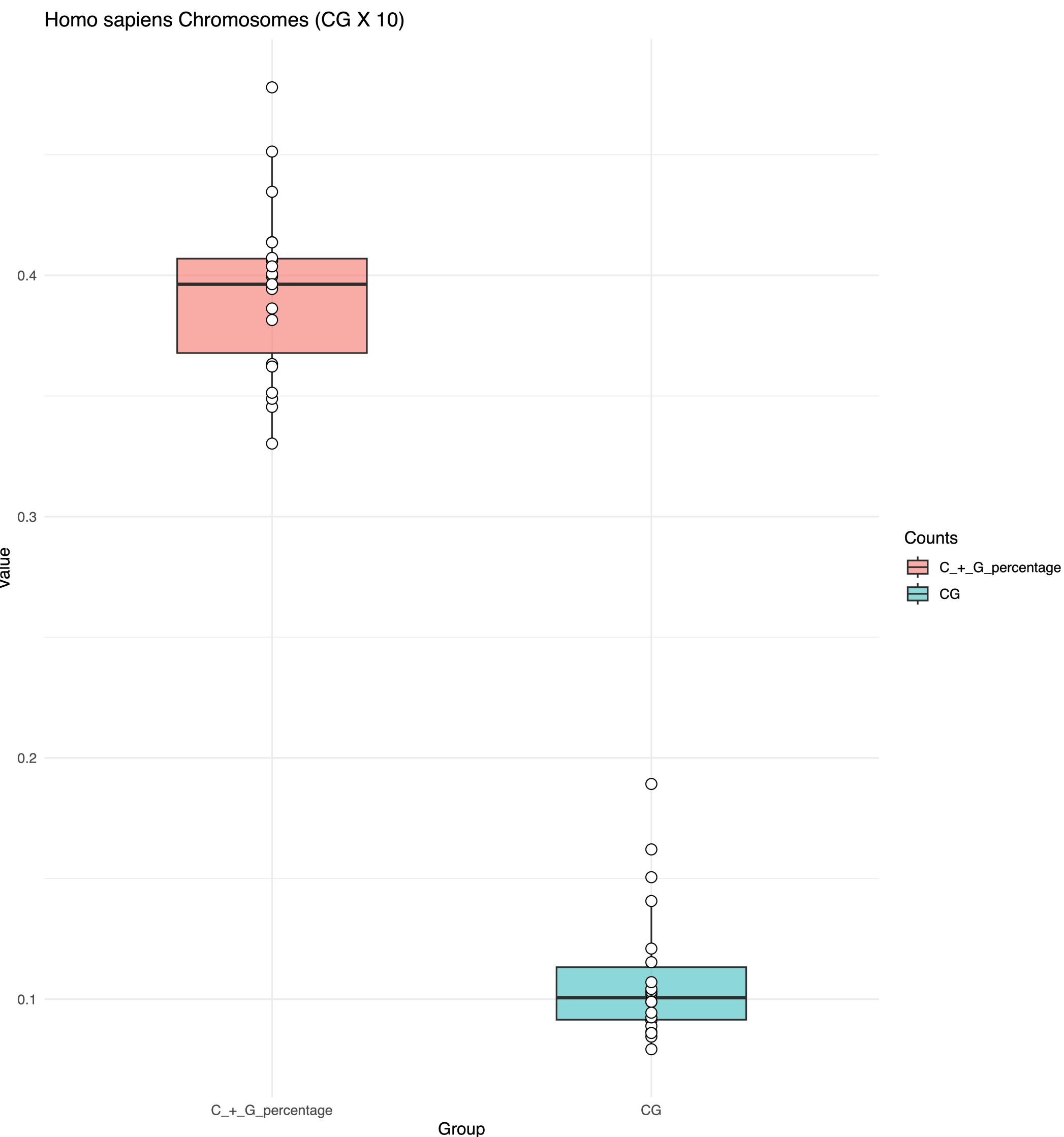
$$f(x) = \begin{cases} \frac{x}{w_A}, & \text{for } 0 \leq x \leq w_A \\ \frac{w_2-x}{w_C} = \frac{w_A+w_C-x}{w_C}, & \text{for } w_A < x \leq w_2 \\ \frac{x-w_2}{w_G} = \frac{x-(w_A+w_C)}{w_G}, & \text{for } w_2 < x \leq w_3 \\ \frac{1-x}{w_T}, & \text{for } w_3 < x \leq 1 \end{cases}$$

$$\begin{aligned} w_1 &= w_A \\ w_2 &= w_A + w_C \\ w_3 &= w_A + w_C + w_G \end{aligned}$$



MENSAJES PARA LLEVAR A CASA

- ▶ Los patrones fractales en el CGR se observan por las frecuencias no homogéneas de los n-omeros.
- ▶ Estos se pueden abordar con la construcción de un formalismo termodinámico a partir de la frecuencias empíricas de los n-omeros
- ▶ Podemos detectar formalmente transiciones de fase que señalan una frontera entre organismo uni y pluricelulares



0

