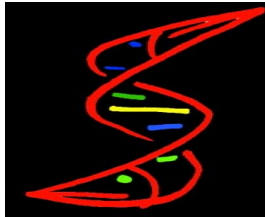


USO DE CODONES SINONIMOS



RSCU (Relative Synonymous codon usage)

$$\text{RSCU}_i = \text{Obs}_i / \text{Exp}_i$$

$$\text{Exp}_i = \sum a_i / \sum \text{syn}_i$$

Example 1:

UUU = 75
<u>UUC = 25</u>
Phe = 100

$$\text{Exp}_{\text{UUU}} = 100 / 2 = 50$$

$$\text{RSCU}_{\text{UUU}} = 75 / 50 = 1.50$$

$$\text{Exp}_{\text{UUC}} = 100 / 2 = 50$$

$$\text{RSCU}_{\text{UUC}} = 25 / 50 = 0.50$$

$$1.50 + 0.50 = 2.0$$



Example 2: GUA = 50
 GUU = 100
 GUC = 20
 GUG = 30
 Val = 200

$$\text{Exp}_{\text{GUA}} = 200/4 = 50$$

$$\text{RSCU}_{\text{GUA}} = 50/50 = 1.00$$

$$\text{Exp}_{\text{GUU}} = 200/4 = 50$$

$$\text{RSCU}_{\text{GUU}} = 100/50 = 2.00$$

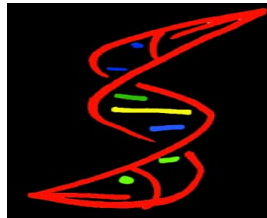
$$\text{Exp}_{\text{GUC}} = 200/4 = 50$$

$$\text{RSCU}_{\text{GUC}} = 20/50 = 0.40$$

$$\text{Exp}_{\text{GUG}} = 200/4 = 50$$

$$\text{RSCU}_{\text{GUG}} = 30/50 = 0.60$$

$$\boxed{1.00+2.00+0.40+0.60 = 4.00}$$



ANALISIS DE CORRESPONDENCIA (COA)

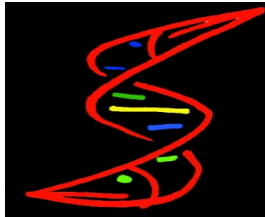
- **El propósito de la estadística ha sido definido como “resumir, simplificar y eventualmente explicar” (Greenacre 1984).**
- **El uso de codones es, por naturaleza, multivariado, y es por lo tanto necesario analizarlo usando técnicas estadísticas multivariadas.**
- **El COA es uno entre varios análisis multivariados.**



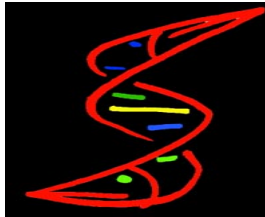
- **Los análisis multivariados se usan para simplificar matrices rectangulares en las cuales (para nuestro propósito) las columnas representan alguna medida de uso de codones o aminoácidos, y las filas representan genes individuales.**



- Dicho en forma **extremadamente simplificada**:
- Cada gen puede ser representado en un **espacio multidimensional** por un vector.
- El espacio multidimensional es **Euclideo**, siendo cada eje **ortogonal**.
- Las distancias son definidas en filas y/o columnas, y estas distancias son aproximadas por distancias **Euclidianas** en una representación de **pocas dimensiones** de la tabla.



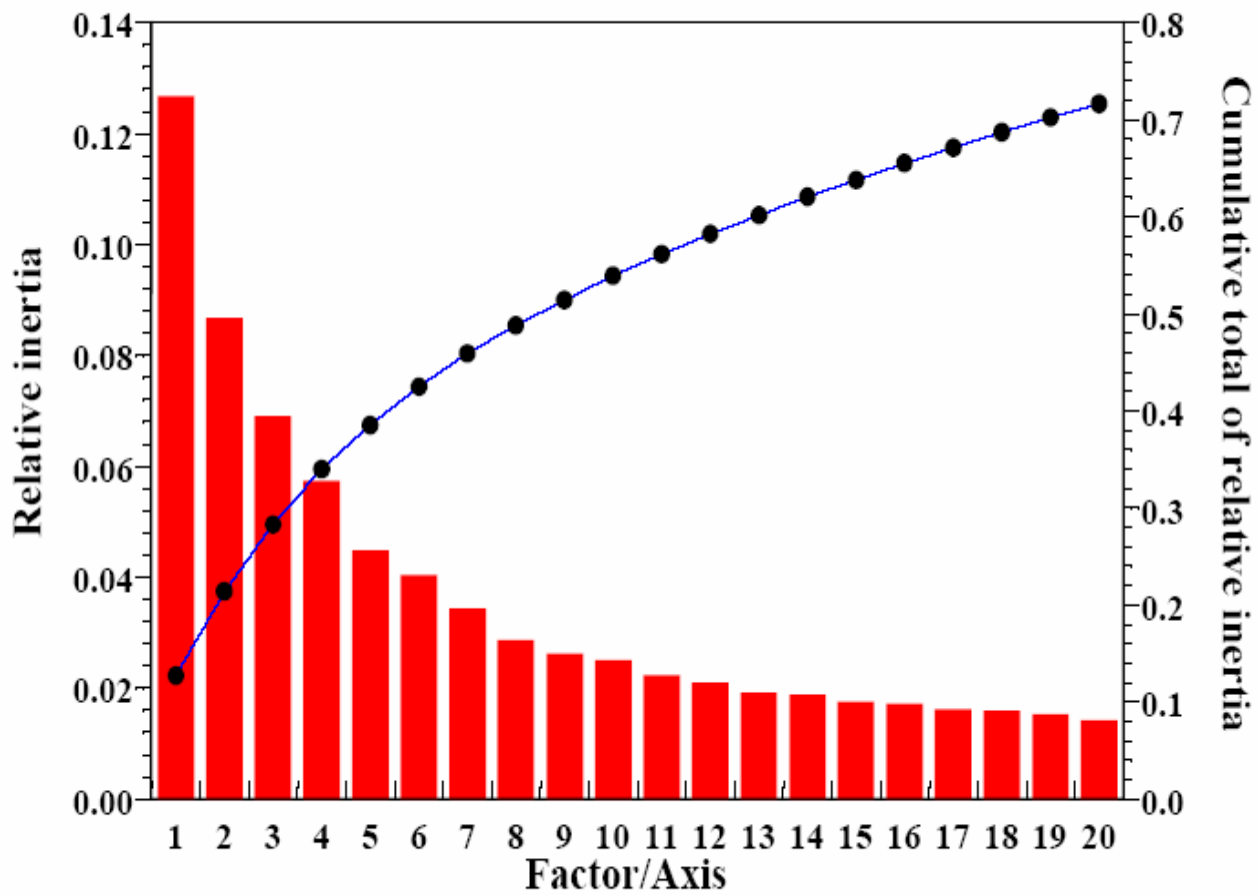
- **El objetivo del COA es identificar el “subespacio” de pocas dimensiones que mejor representa los puntos (genes).**



ANALISIS DE CORRESPONDENCIA (COA)

- **Es una herramienta poderosa para “aislar” tendencias, pero su mayor debilidad consiste en que no brinda claves para la interpretación de estas tendencias.**
- **La interpretación corre por cuenta del investigador.**





ORGANISMOS UNICELULARES

Si las terceras posiciones son selectivamente neutras (sinónimas**), entonces todos los tripletes que codifican para el mismo aa, estarían presentes en la misma frecuencia si analizamos un número suficientemente grande de genes.**



- Pero se vio que existe una gran variación interespecífica, lo que llevó a postular la "genome hypothesis" (Grantham et al., 1980).
- Esta variación puede ser explicada por la influencia de distintos sesgos mutacionales.

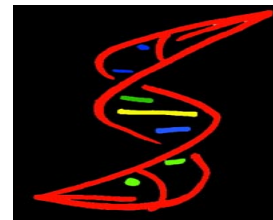
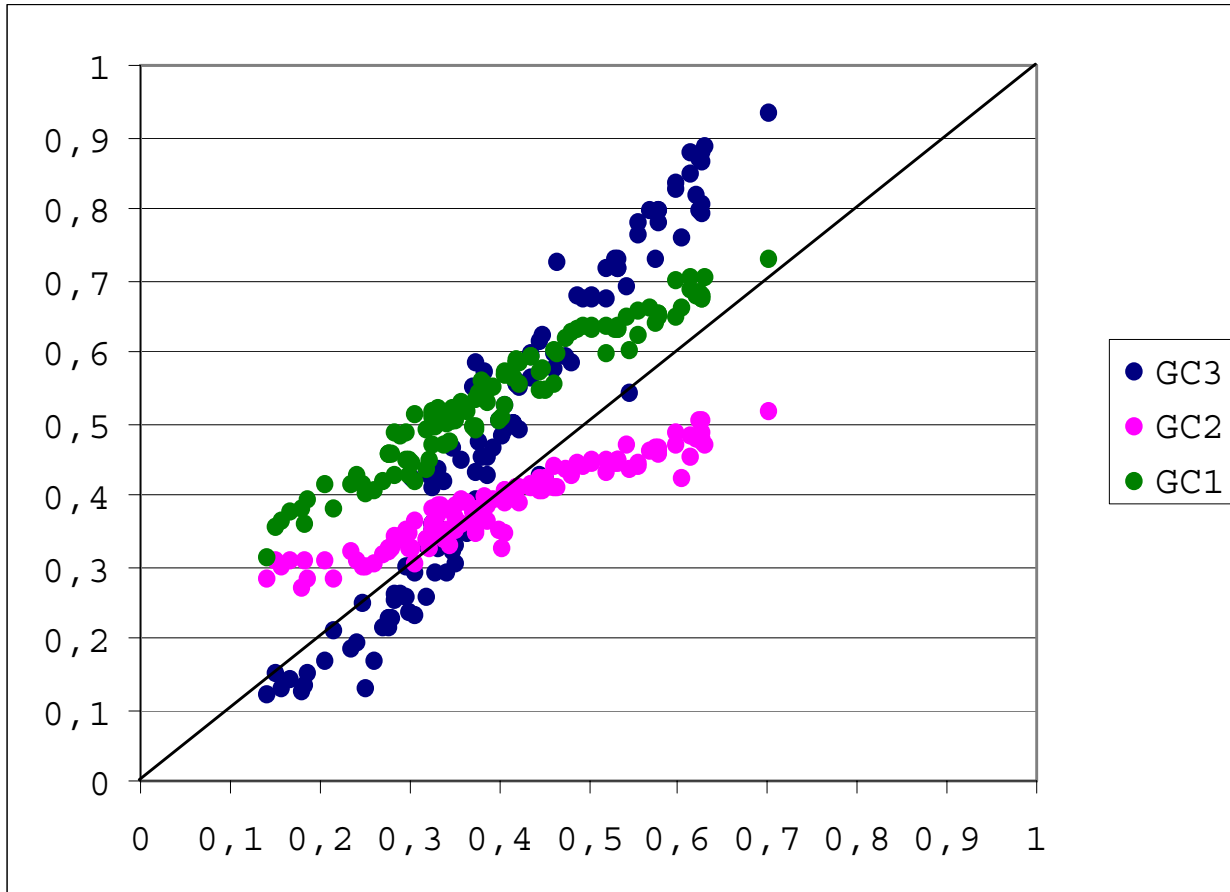


Propiedades composicionales de algunas bacterias

	<i>C. perf.</i>	<i>E. coli</i>	<i>M. tub</i>
GC genoma	0.29	0.50	0.66
GC 1	0.41	0.57	0.68
GC 2	0.31	0.41	0.50
GC 3	0.16	0.53	0.79



Correlaciones composicionales entre las tres posiciones de los codones y el GC genómico (regiones intergénicas)

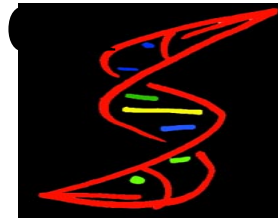


C. perf.

E. coli

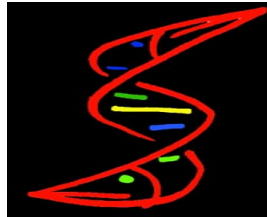
M. tub

• GC genoma	0.29	0.50	0.66
• GC 1	0.41	0.57	0.68
• GC 2	0.31	0.41	0.50
• GC 3	0.16	0.53	0.79
• Ala GCA	0.37	0.23	0.10
• GCC	0.08	0.26	0.45
• GCG	0.02	0.32	0.37
• GCU	0.54	0.19	0.08
• Phe UUC	0.20	0.42	0.79
• UUU	0.80	0.58	



Variabilidad intragenómica

- **Entre un grupo de codones sinónimos reconocidos por varios ARNt, los reconocidos por el ARNt más abundante son usados más frecuentemente (adaptación del uso de codones al pool de ARNt)**
- **Entre los codones reconocidos por el mismo ARNt, los que tienen puentes de H naturales W-C con la posición de tambaleo del anticodón, son generalmente preferidos.**
- **El sesgo es más extremo en los genes de más alta expresión (selección para velocidad, concepto de **codones mayores**).**



Uso de codones en genes de baja expresión

- **Teoría de regulación de la expresión:** los codones raros son usados como un mecanismo para mantener su expresión bajos niveles.
- **Teoría de la selección-mutación-deriva:** el uso de codones en una población finita resulta del balance entre la selección natural y la mutación junto con la deriva, permitiendo que persistan en el genoma codones no óptimos.



Selección natural

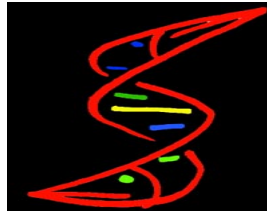
Tres aspectos de la traducción podrían ser afectados por el uso de codones sinónimos:

- a) la tasa de elongación (velocidad)**
- b) el costo del “proofreading”**
- c) la fidelidad en la traducción**



a) la tasa de elongación

- 1) Los genes de más alta expresión presentan un uso de codones más sesgado.
- 2) Muestran un incremento significativo de codones mayores.
- 3) Usualmente estos codones aparean perfectamente con el ARNt isoceptor más abundante.
- 4) La tasa de sustituciones sinónimas es más baja en los genes de alta expresión, y viceversa.
- 5) Evidencias experimentales: **En *E. coli*, la traducción de los codones mayores ocurre 3 a 6 veces más rápido que la de los menores, permitiendo un mejor uso de los ribosomas.**



b) el costo del “proofreading”

- **Los codones mayores pueden reducir el costo energético de rechazar un ARNt equivocado, lo que cuesta un GTP por reacción**



c) la fidelidad en la traducción

- **Las propiedades críticas de las proteínas dependen de aas particulares en posiciones específicas de sus estructuras. La traducción “fiel” reduce los costos de producir proteínas no funcionales por incorporaciones erróneas y/o errores en la procesividad (“frameshifting” y/o terminaciones prematuras).**



- **En *E. coli*, la traducción de los codones mayores es 10x más “fiel” que la de los codones menores, permitiendo también un uso más eficiente de los ribosomas.**



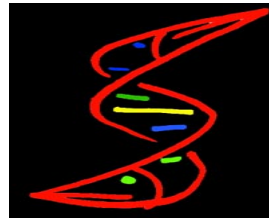
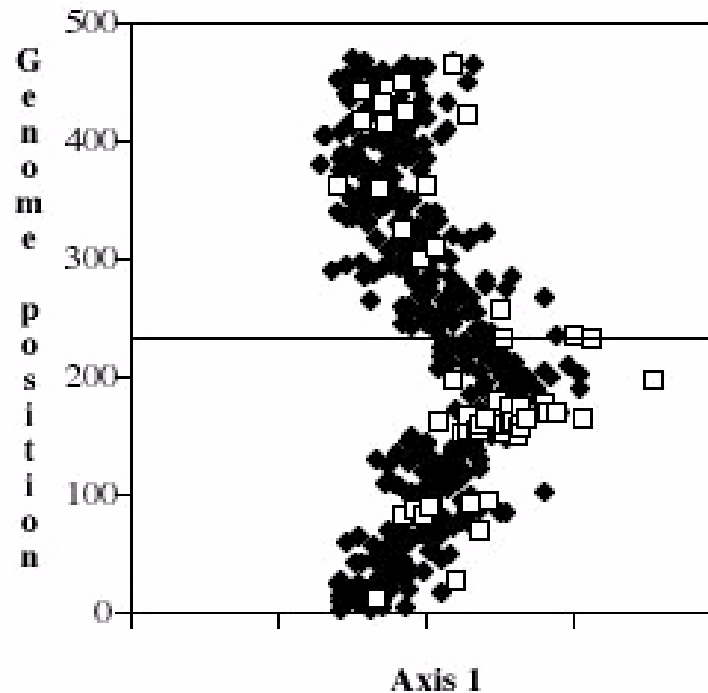
- Por lo tanto, la hipótesis más aceptada para microorganismos, establece que el uso de codones sinónimos es el resultado del balance entre el **sesgo mutacional** (que actúa sobre todo el genoma) y la **selección natural** (operativa a nivel de la traducción).



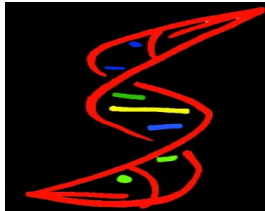
En los últimos años, la disponibilidad de genomas completos de procariotas permitió detectar nuevos factores que influyen en el uso de codones sinónimos, por ejemplo:

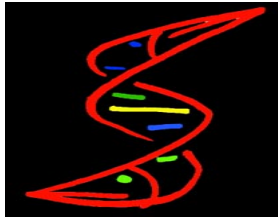
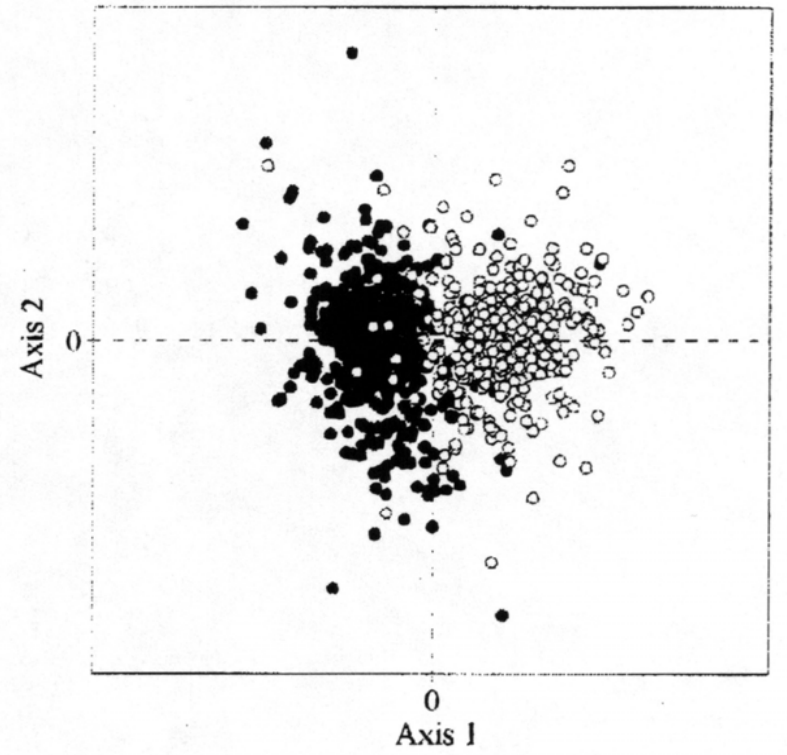
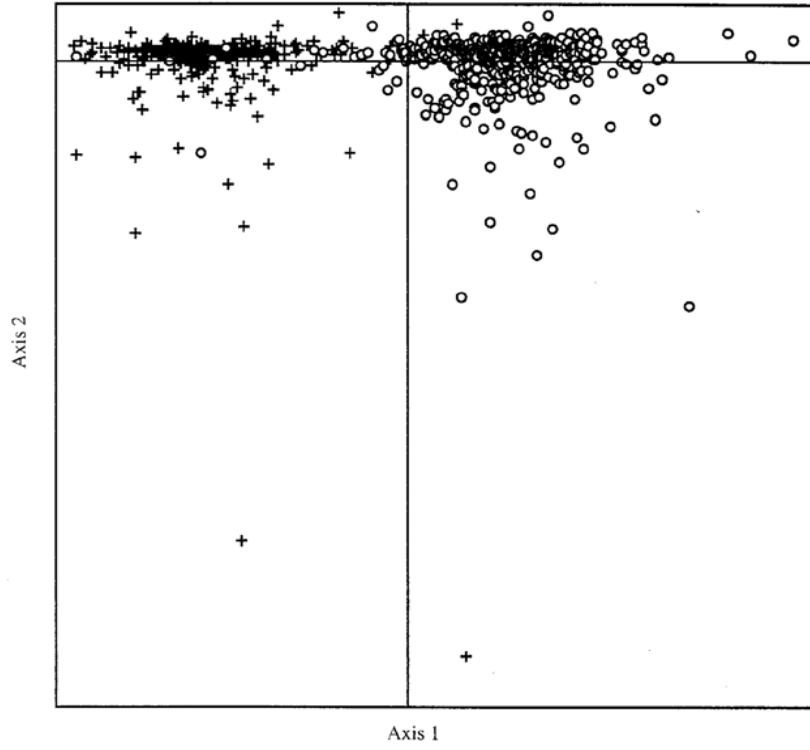


1) La localización en el genoma de cada secuencia determina el GC3, y por lo tanto el uso de codones sinónimos en *Mycoplasma genitalium*.



2) En especies como *Borrellia burgdorferi*, *Treponema pallidum* y *Chlamydia trachomatis*, la localización de los genes en la hebra “leading” o “lagging” de la replicación es el factor principal en el uso de codones sinónimos.





- 3) En **Mycobacteria** y *Chlamydia trachomatis* la hidropatía de cada proteína codificada, está entre los factores de más influencia en el uso de codones.
- 4) Los procariotas **mesofílicos** y **termofílicos** difieren en el uso de codones a nivel de **Arg**: los primeros prefieren **CGN** y los segundos **AGR**.

