

# Análisis básico

Bioinformática  
Máster en Biotecnología

# Estructura composicional

## El contenido en G+C (%G+C)

$$\%(G+C) = 100 * (\text{frec}(G) + \text{frec}(C)) / (\text{frec}(A) + \text{frec}(C) + \text{frec}(G) + \text{frec}(T))$$

**Parches cortos (entre 100 bp y varios kb) ricos en %G+C pueden tener significado biológico**

- Los exones y mas marcadamente la región promotora se suelen desmarcar de su entorno mediante un G+C mas alto
- Las retro-transposones Alus son ricas en G+C
- Un región rica en G+C en una Bacteria rica en A+T puede indicar un evento de transferencia horizontal
- Isocoras en los genomas de muchos mamíferos

# Composición en dinucleótidos

**La composición de dinucleótidos se calcula en ventanas solapantes**

Por ejemplo la secuencia ATCGATG

Contiene los dinucleótidos 'AT', 'TC', 'CG', 'GA', 'AT' y 'TG'

Para obtener información acerca de que si el número observado de dinucleótidos en una secuencia se desvía de lo esperado, calculamos la ratio O/E

Ratio O/E = frecuencia observada / frecuencia esperada

La frecuencia esperada de un dinucleótido 'NM' se calcula:  $(\#N * \#M) / (\text{longitud de secuencia en nt})^2$

**O/E >> 1: mas dinucleótidos de lo esperado**

**O/E << 1: menos dinucleótidos de lo esperado**

**En mamíferos observaremos mucho menos dinucleótidos CG de lo esperado.**

**¿Por qué?**

# Las isocoras

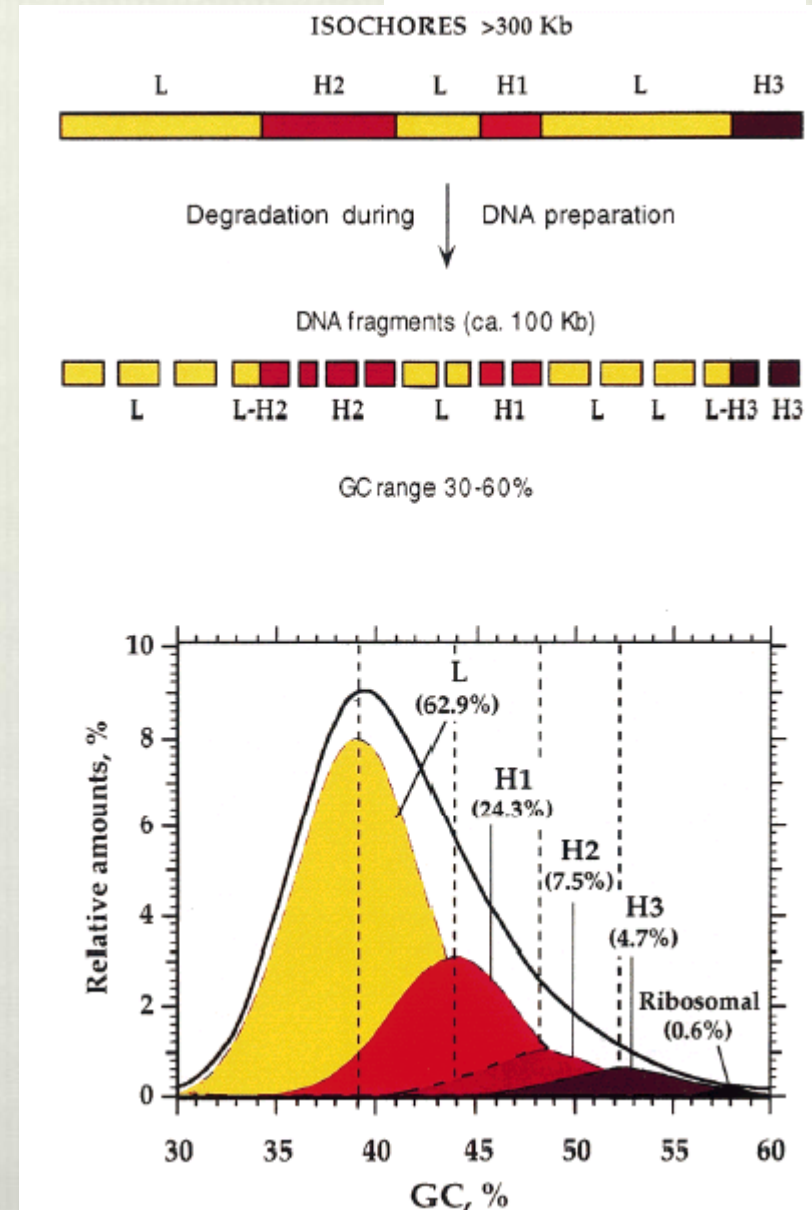
G. Bernardi / *Gene* 241 (2000) 3–17

**Estructura composicional: propiedades derivadas de la secuencia. No se refiere a propiedades citogenéticas**

## Isocoras:

tramos largos de ADN ( $\gg 300$  kb)  
'relativamente' homogéneos  
composicionalmente.

Fig. 1. (Top) Scheme of the isochore organization of the human genome. This genome, which is typical of the genome of most mammals, is a mosaic of large ( $\gg 300$  kb, on average) DNA segments, the isochores, which are compositionally homogeneous (above a size of 3 kb) and can be partitioned into a number of families. Isochores are degraded during routine DNA preparations to fragments of approx. 100 kb in size. The GC-range of the isochores from the human genome is 30–60% (from Bernardi, 1995). (Bottom) The CsCl profile of human DNA is resolved into its major DNA components, namely the families of DNA fragments derived from isochore families L (i.e., L1+L2), H1, H2, H3. Modal GC levels of isochore families are indicated on the abscissa (broken vertical lines). The relative amounts of major DNA components are indicated. Satellite DNAs are not represented (from Zoubak et al., 1996).



# Predicción de isocoras (chr22)

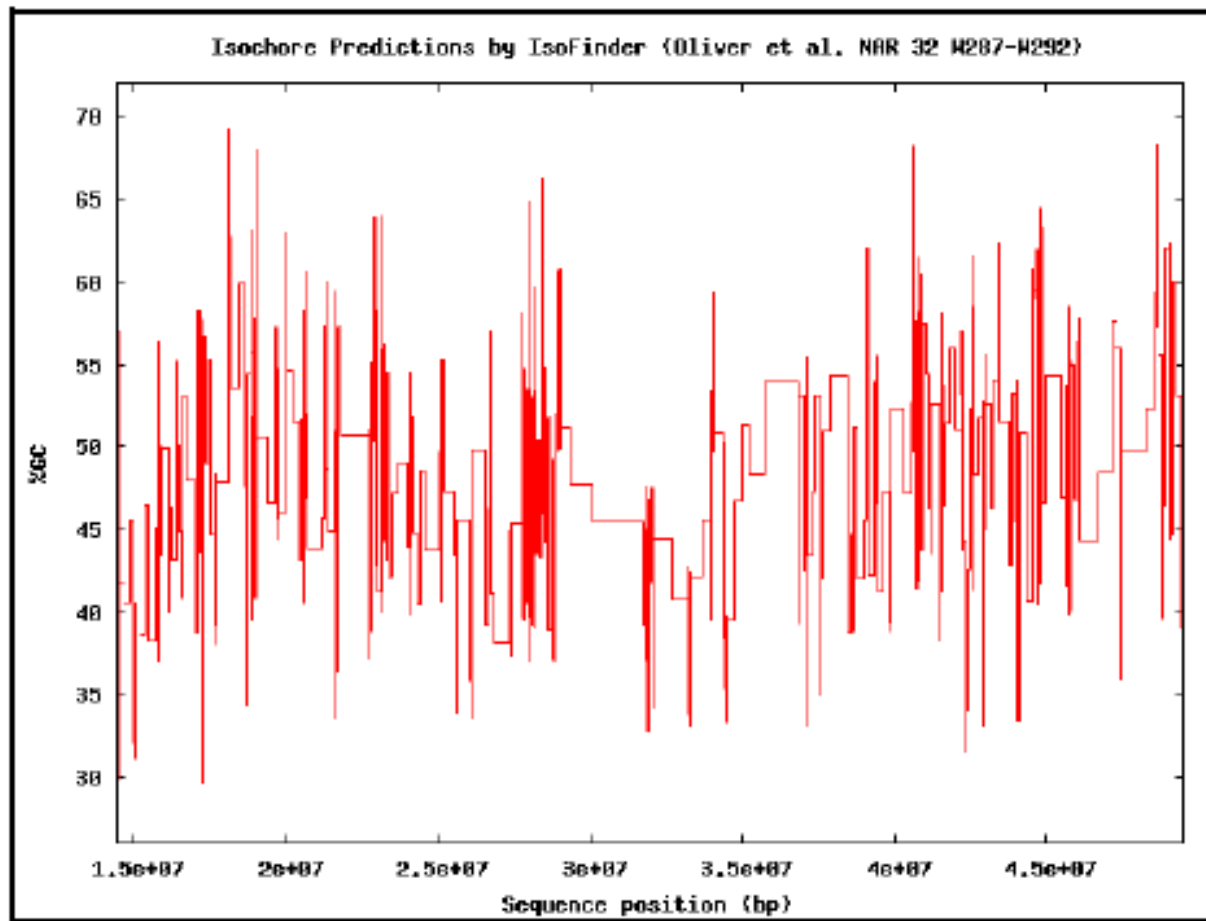


Figura 1-5: Representación grafica de la segmentación del cromosoma 22 generada mediante la interfaz web del programa IsoFinder (Oliver *et al.*, 2004). Se puede apreciar la variación pronunciada del contenido en G+C entre aproximadamente el 35 y el 65%.

# Relevancia biológica de las isocoras

- Correlación con las bandas cromosómicas G(rico en A+T)/R (rico en G+C)
- Densidad de genes: mas alto en isocoras H
- Densidad de islas CpG y transposones (Alus, LINEs)
- Sesgo del uso de codones
- Sesgo mutacional
- Replicación (momento de replicación depende del G+C)
- Frecuencia de recombinación
- Estructura génica (longitudes de genes e intrones)
- Observan principalmente en animales de sangre caliente
- ¿Patrones de metilación?
- ¿Especificidad tisular de la expresión génica?