



# ***A matemática do DNA***

**Vera Afreixo**

**Departamento de Matemática  
Universidade de Aveiro  
6 de Junho de 2009**

**Encontro de antigos alunos 2009**

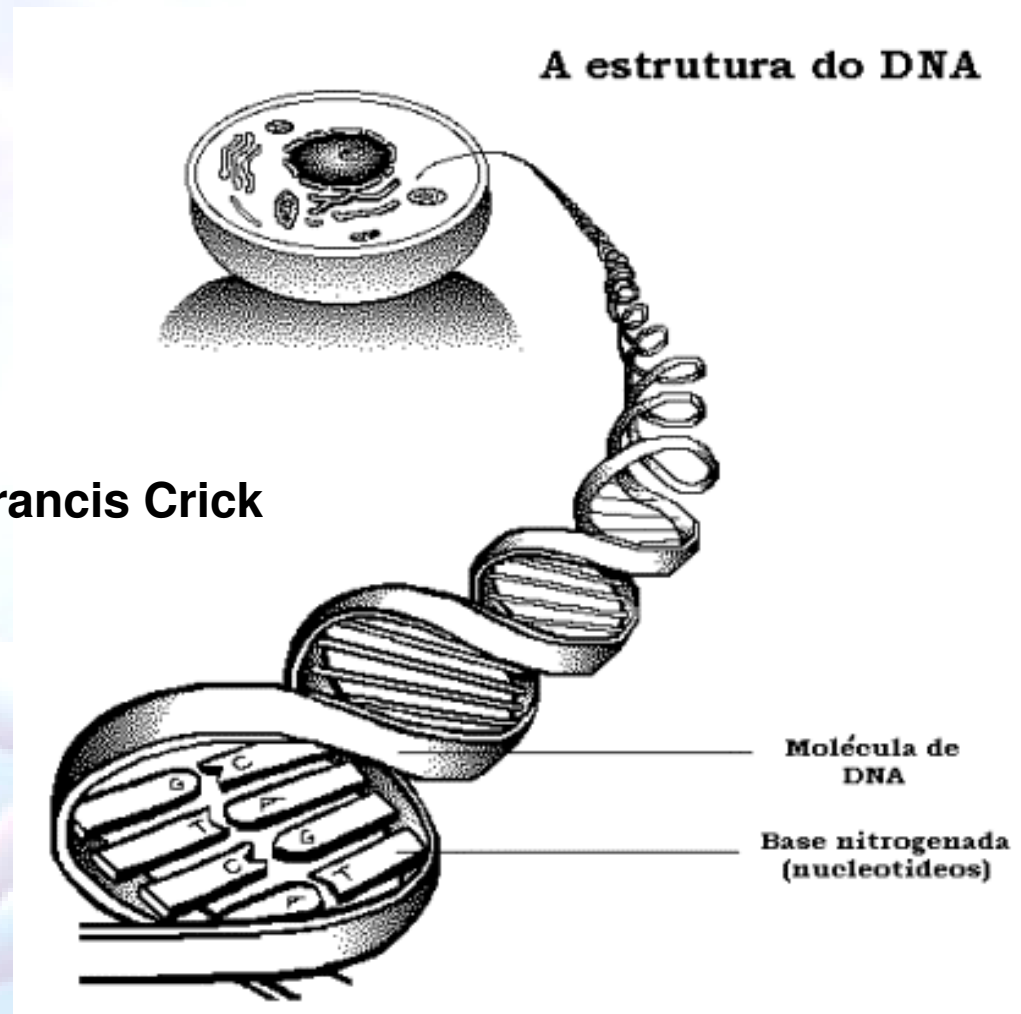
# *Conteúdo*

- Introdução
  - DNA
- Estado da arte
  - matemática aplicada ao DNA
- Modelo matemático
  - evolução das espécies
- Conclusões

# Introdução DNA

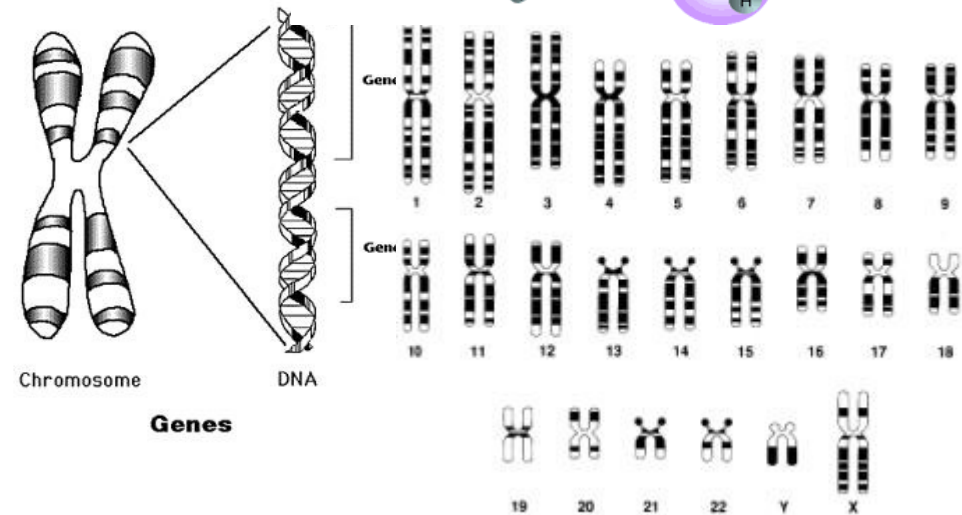
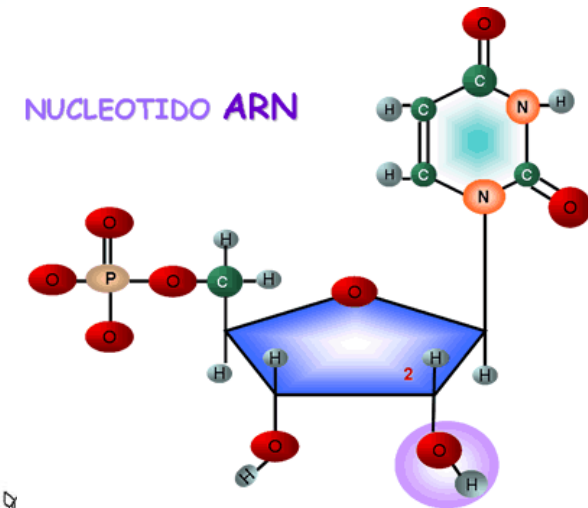
## O DNA

1953 James Watson e Francis Crick



# Introdução DNA

- Nucleótidos **ACGT**
- Cromossomas
  - Região codificante (genes)
  - Região não codificante
- Genoma completo



# *Introdução*

## *DNA*

- O DNA é essencialmente constituído por 4 moléculas diferentes **ACGT**
- Cada ser vivo tem um DNA específico
- Através do DNA qualquer célula reproduz-se e fabrica as proteínas necessárias à sua sobrevivência.

**Actualmente em páginas públicas encontram-se genomas de várias espécies completamente sequenciados.**

# Introdução DNA

## Resultado da sequenciação do DNA

AAACAATCTCGATTCTAAATTGAAACGAACGCAGCATTTCAGGGACTGGATGAGGA  
GCTTACGGTTTTTTACAGAATCATCAATATCTTGGAAGAAAAAGAATGTTAAGAAAT  
AACAAAACAATAATTATTAAGTACTTTCTTAATCTCATTAAATGGAGCTTTCTTGGTTCT  
TGGACTTTTATTCATGGGATTTGGTGCATGGCTCTTATTAGATAGAAATAATTTTTTA  
ACAGCTTTTGATGAAAATAATCACTTCATAGTACCTATTTCTCAAATTTTGATTGGAA  
TGGGATCTTCTACTGTTCTTTTTTGTCTATTGGGTTATATAGGAATTCACAACGAAAT  
CAGATGGCTCCTAATTGTGTATGCAGTATTGATAACATGGACCTTTGCTGTTTCAGGT  
TGTACTTTCAGCATTTCATCATCACAAGAAAGAGGAGGTTTCAGCAACTATGGCATGA  
CAAAATTGATTTTGTCAATTTCTGAGTATGGATCTAAAGATAAGCCTGAAGATATAAC  
CAAGTGGACTATTCTGAATGCCTTACAGAAAACATTACAGTGTTGTGGCCAACATAA  
TTACACAGACTGGATAAAGAATAAGAACAAGAAAATTCAGGACAGGTGCCATGTT  
CTTGCACAAAGTCAACTTTAAGAAAATGGTTTTTGTGATGAGCCACTGAATGCAACTT  
ACCTTGAGGGTTGTGAAAATAAAATCAGTGCATGGTATAATGTTAATGTGTTAACCT  
TAATCGGAATTAACTTTGGACTTTTAACTTCAGAGGTTTTCCAAGTCTCATTAAACAGT  
TTGTTTCTTCAAAAACATCAAGAATATAATCCATGCAGAAATGTGACCTTTGGATTTC  
AATTTGTTTCAGAAGAAACCAGTTAATTCTTAAAAAATCACATTATATTATTTTATTCCA  
AAAACGTTTTTGAATTGTATTAATAACAGATGTATTCAAATTATTAATTATATGTAATA  
TTGAATTACTGAATACTATAATAAAACATTCATGAAATTATAAAAAAAAAAAAAAAAAA

# ***Estado da arte***

## ***matemática aplicada ao DNA***

- A matemática desempenha um papel fundamental
  - na tarefa de entender como o DNA funciona e como ele se replica
    - transformar microorganismos para os converter em autênticas fábricas que produzem substâncias úteis, como p.e. a insulina
  - na localização de genes ou outro tipo de regiões específicas nas longas sequências de DNA
  - na compressão das longas sequências de nucleótidos
  - na construção de árvores filogenéticas
  - ...

# ***Estado da arte***

## ***matemática aplicada ao DNA***

- As sequências genéticas têm natureza simbólica o que não permite aplicar directamente metodologias de análise de sequências numéricas
  - Mapeamento directo de símbolos em números
  - Mapeamento de símbolos em números que traduzem estrutura da sequências
    - Autocorrelação da sequência (função de autocorrelação simbólica)
    - Distância entre nucleótidos iguais



# Modelo matemático evolução das espécies

AAACCCGTGTCAGTT

## Distâncias entre nucleótidos

$d = (1, 1, 9, 1, 1, 5, 2, 2, 4, 4, 8, 4, 9, 1, 8)$

$d^A = (1, 1, 9, 4)$

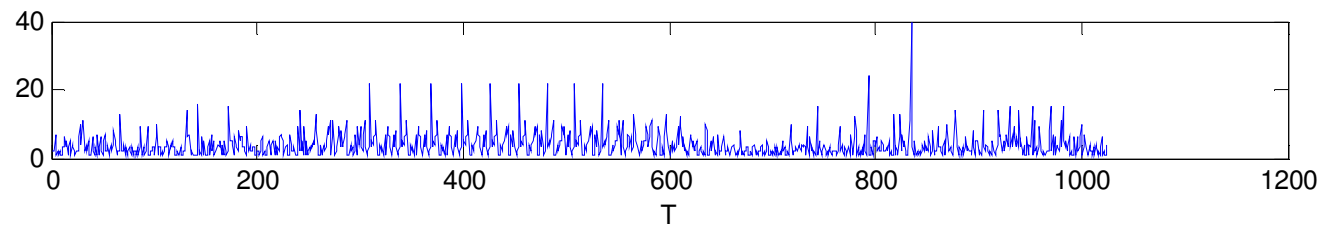
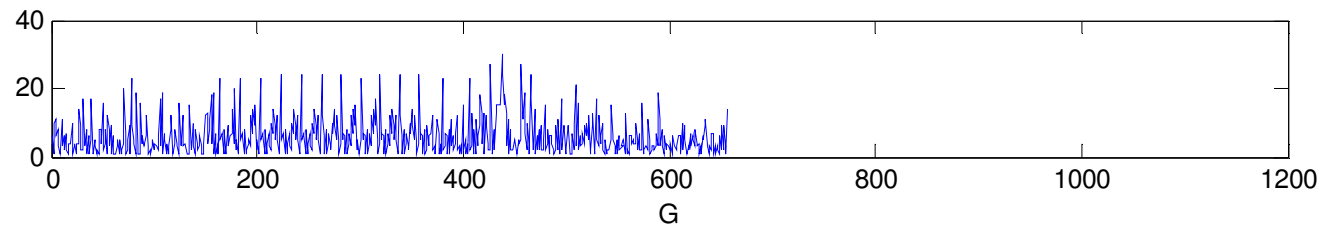
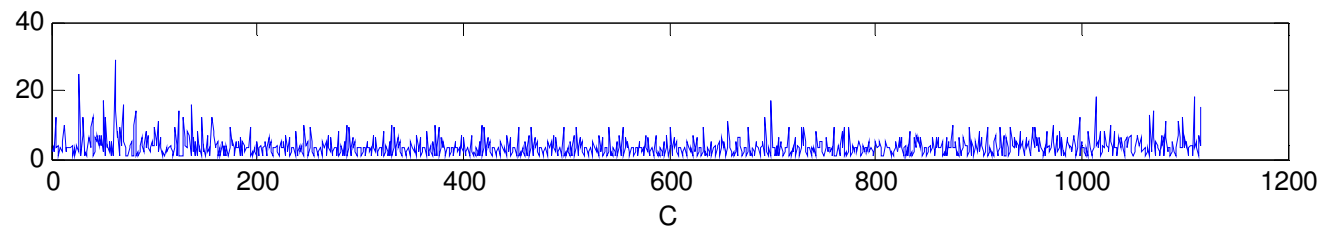
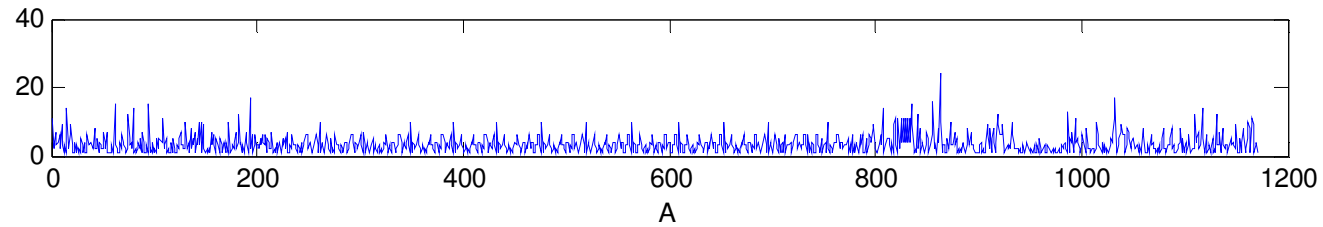
$d^C = (1, 1, 5, 8)$

$d^T = (2, 4, 1, 8)$

$d^G = (2, 4, 9)$

# *Modelo matemático*

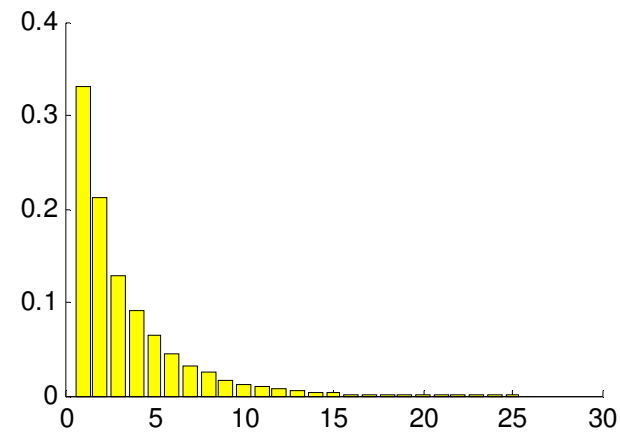
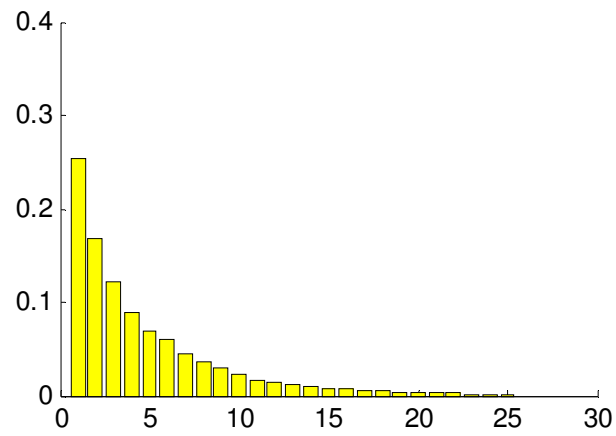
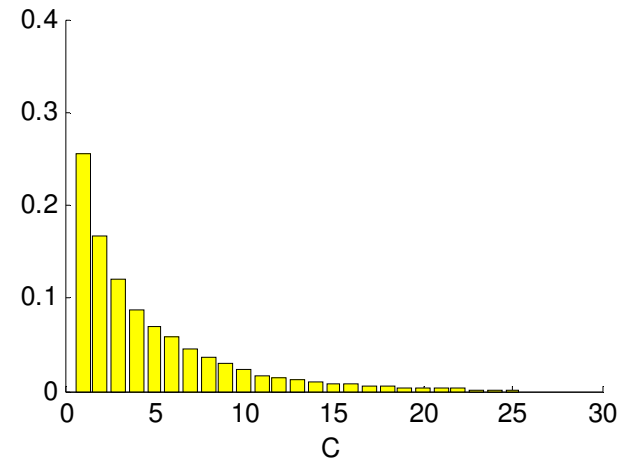
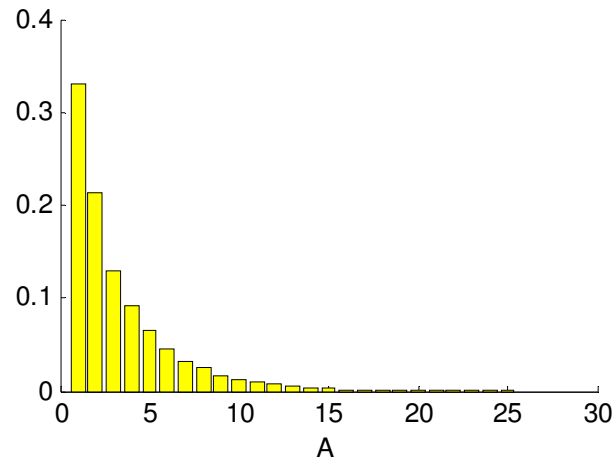
## *evolução das espécies*



Distâncias entre nucleótidos

# Modelo matemático

## evolução das espécies



**Distribuição das distâncias**

# *Modelo matemático*

## *evolução das espécies*

Comportamento: aleatório e independente

$$P(A)=p$$

$$P(D=1)=P(A|A)=p$$

(AA)

$$P(D=2)=P(A^cA|A)=(1-p)p$$

(AXA)

$$P(D=3)=P(A^cA^cA|A)=(1-p)^2p$$

(AXXA)

$$P(D=4)=P(A^cA^cA^cA|A)=(1-p)^3p$$

(AXXXA)

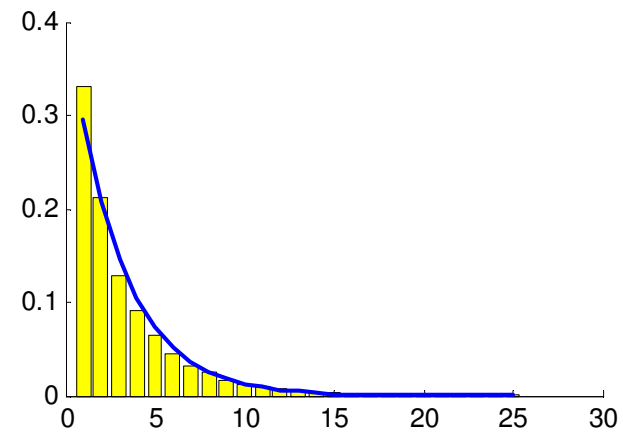
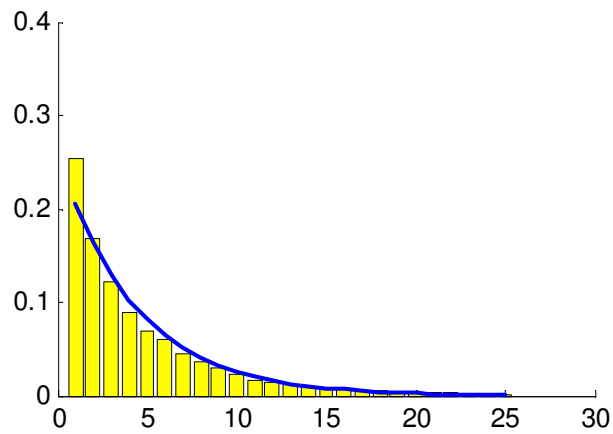
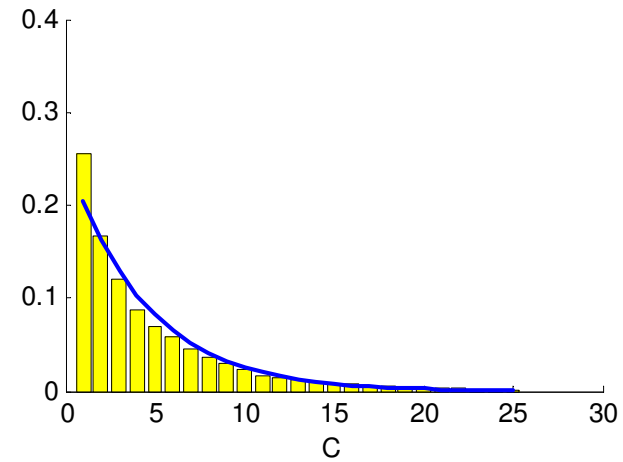
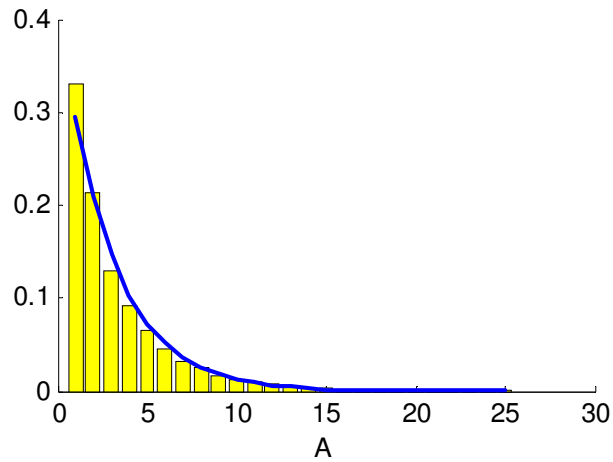
...

...

$$D \sim G(p)$$

# *Modelo matemático*

## *evolução das espécies*



**Distribuição das distâncias**

# ***Modelo matemático evolução das espécies***

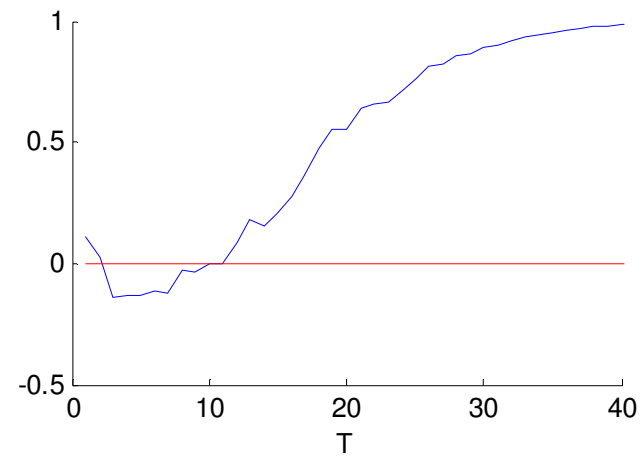
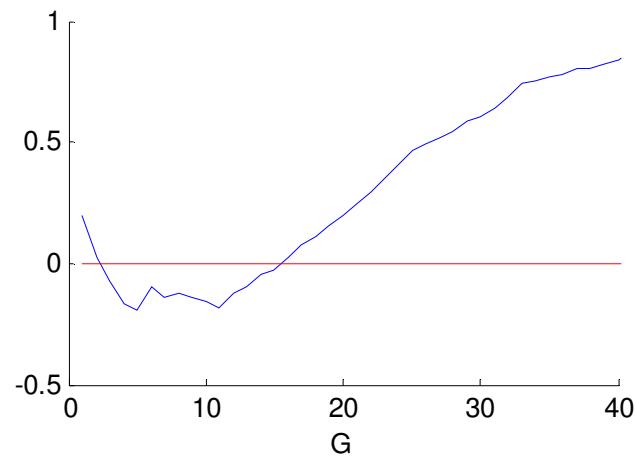
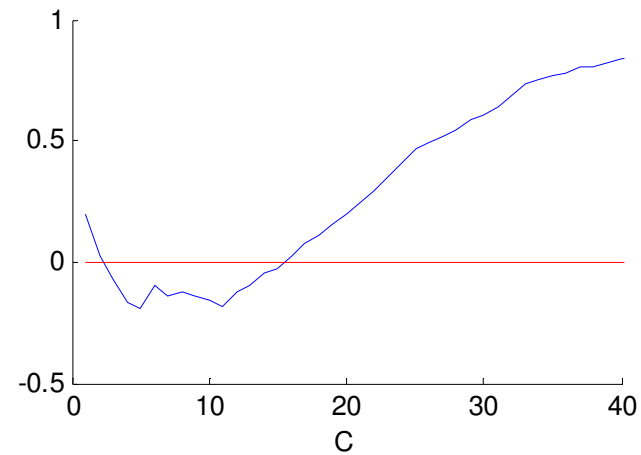
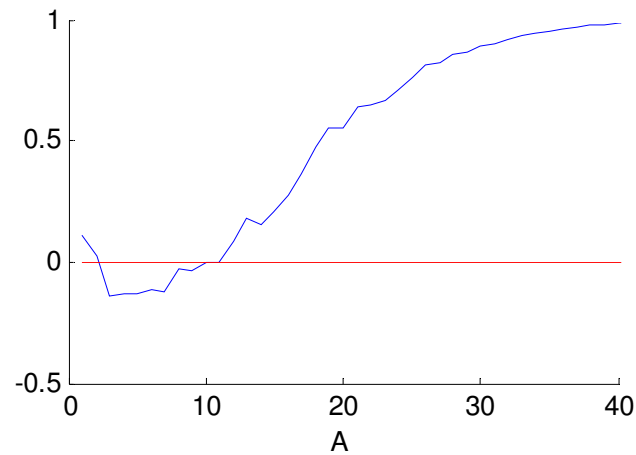
**Erros relativos**

$$(f_o - f_g)/f_o$$

$f_o$  - distribuição observada

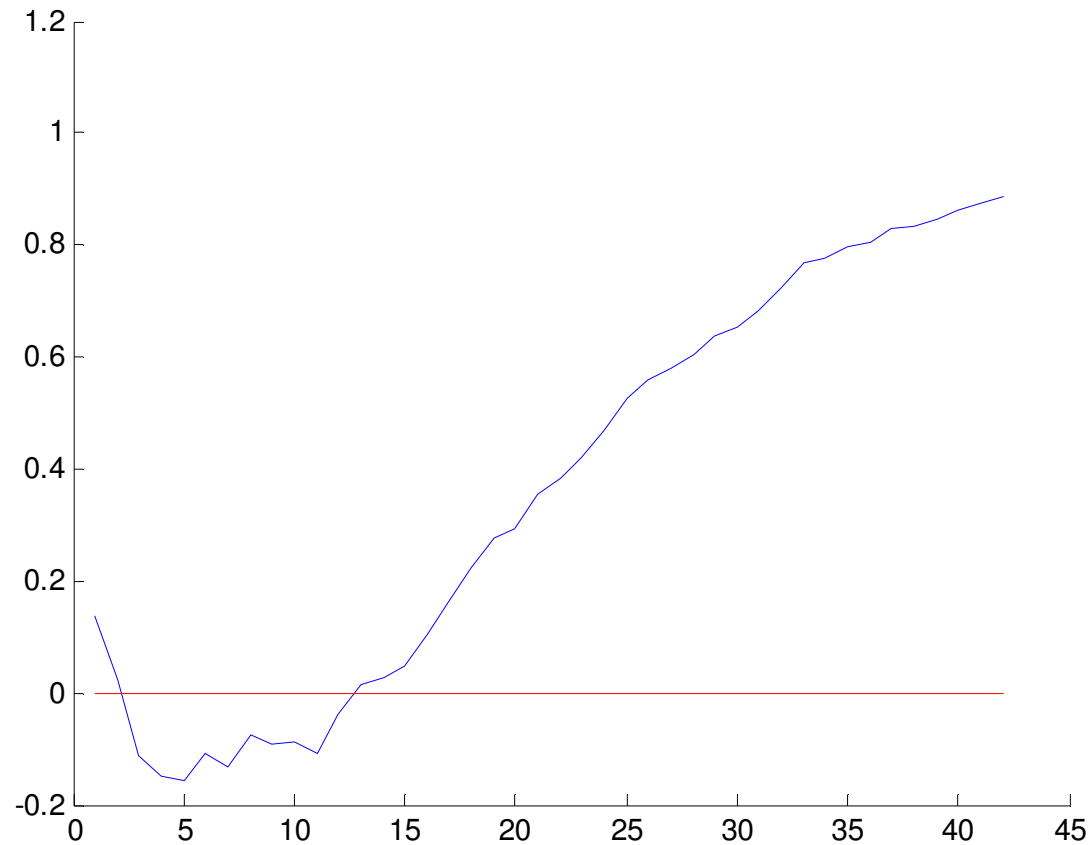
$f_g$  - distribuição geométrica

# Modelo matemático evolução das espécies



**Erros relativos do genoma humano**

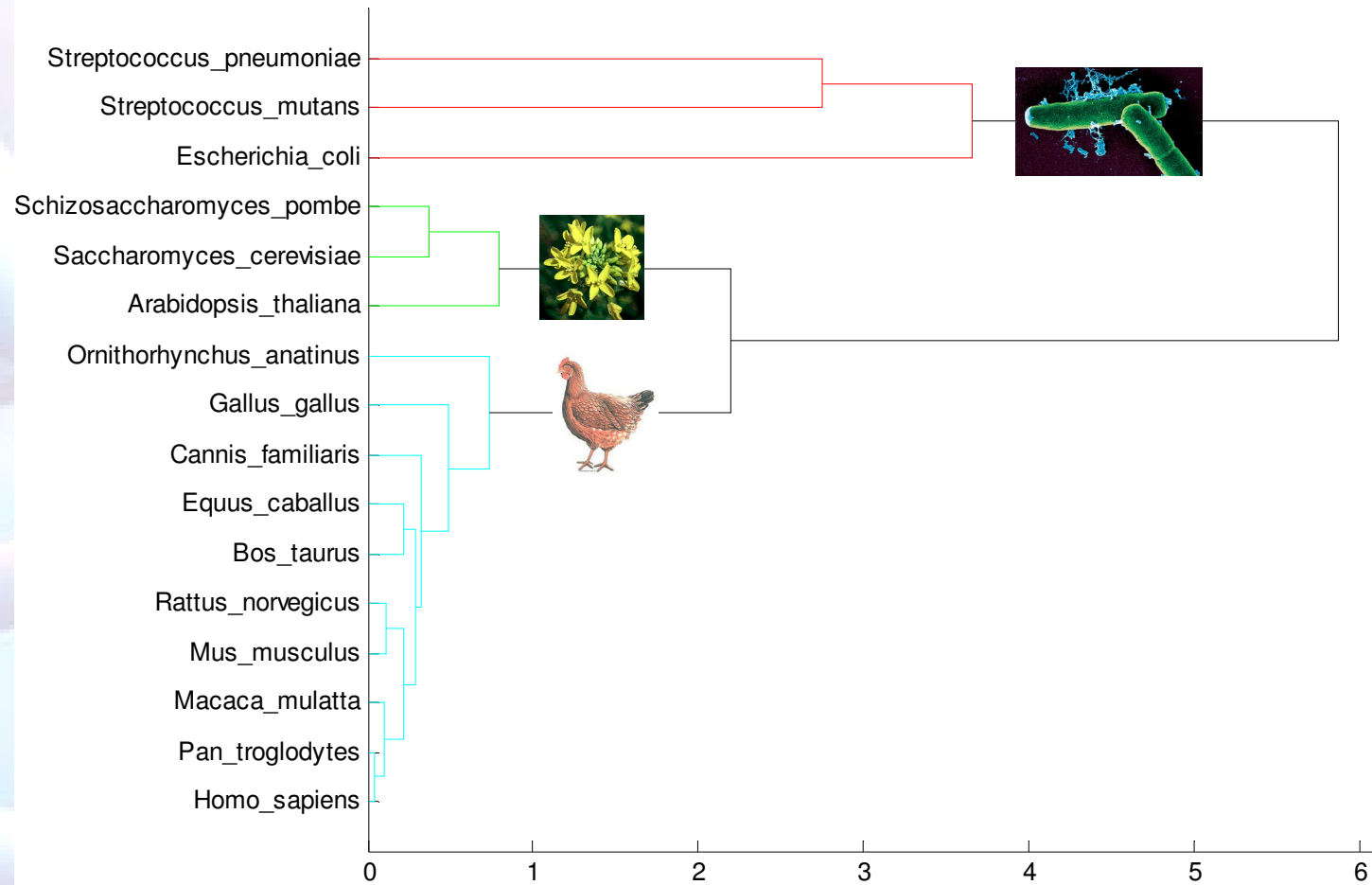
# *Modelo matemático evolução das espécies*



**Erros relativos do genoma humano  
assinatura genómica**

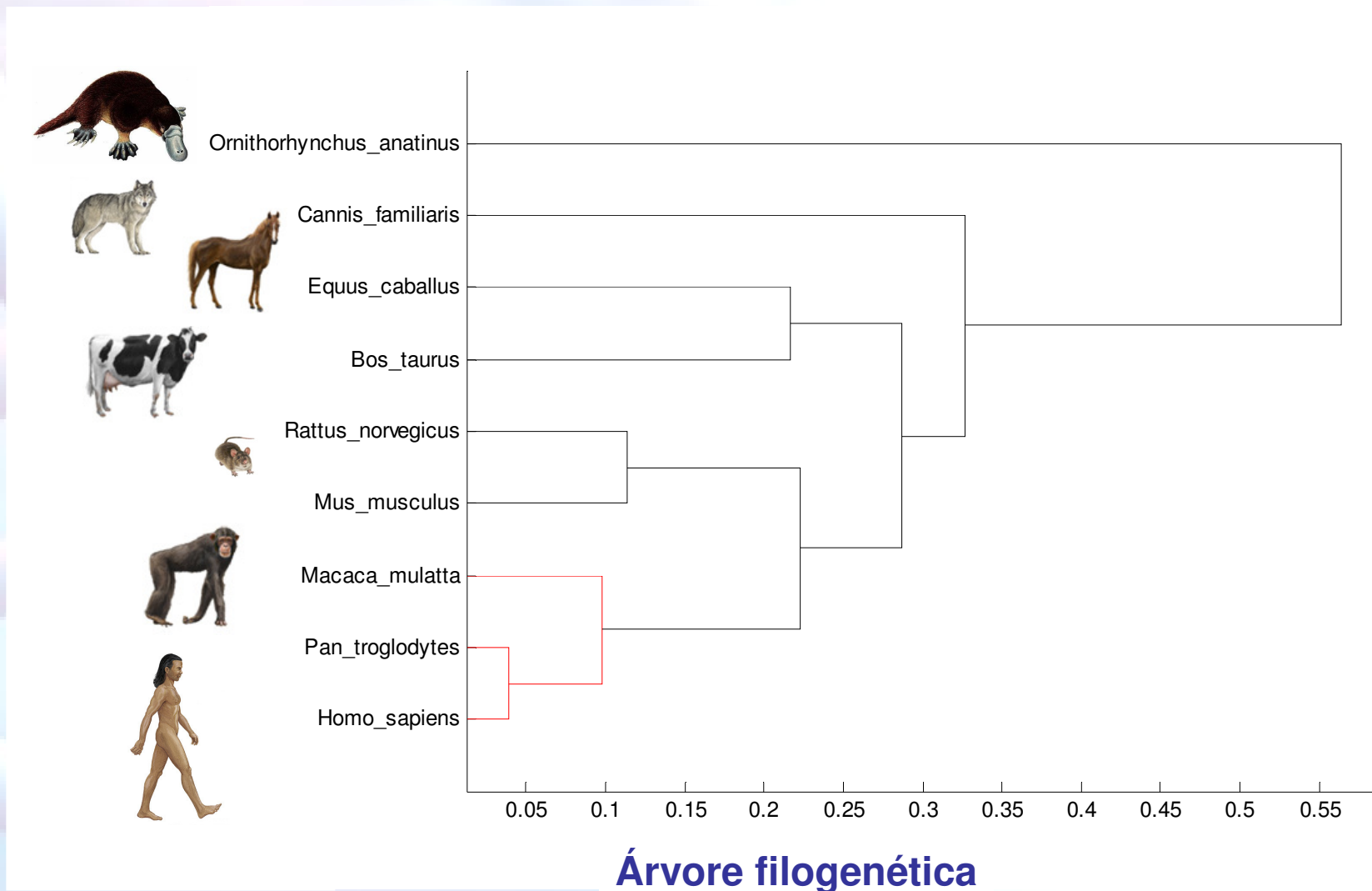


# Modelo matemático evolução das espécies



Árvore filogenética

# Modelo matemático evolução das espécies



# *Conclusões*

- Com este modelo matemático apresentámos uma contribuição para explorar e comparar sequências de DNA
- A matemática do DNA foi, é e será objecto de (muito) trabalho de muitos curiosos