

# Introdução à Biologia Molecular Computacional

História

Cadeias de DNA e de Proteínas

# Meio de 1970 – Primeiros métodos de seqüenciamento rápido de DNA

- Em 1975, F. Sanger, S. Nicklen e A. R. Coulson desenvolveram o primeiro método de sequenciamento rápido de DNA, chamado chain-termination method.
- Em 1976-1977, A. Maxam e W. Gilbert desenvolveram um novo método, modificação química, menos tóxico, que eventualmente se tornou mais popular.

# Até o início da década de 90 ...

Os únicos organismos seqüenciados eram procariotos, que incluem as bactérias, normalmente associados a alguma doença ou de interesse na pesquisa científica.

<http://pt.wikipedia.org/wiki/Procarionte>

A espécie bacteriana [\*Escherichia coli\*](#) se destaca como organismo modelo e como ferramenta biológica para pesquisas científicas (antibióticos).

# Hierarquia da Classificação

([Wikipedia.org/](http://Wikipedia.org/) [LINK](#))

<b>Taxon</b>	<b><u>Humano</u></b>	<b><u>Ervilha</u></b>	<b><u>Amanita</u></b>	<b><u>E. coli</u></b>
<b><u>Domínio</u></b>	<u>Eukaryota</u>	<u>Eukaryota</u>	<u>Eukaryota</u>	<u>Bacteria</u>
<b><u>Reino</u></b>	<u>Animalia</u>	<u>Plantae</u>	<u>Fungi</u>	<u>Monera</u>
<b><u>Phylum</u> ou <u>Divisão</u></b>	<u>Chordata</u>	<u>Magnoliophyta</u>	<u>Basidiomycota</u>	<u>Proteobacteria</u>
Subphylum ou subdivisão	<u>Vertebrata</u>	<u>Magnoliophytina</u>	<u>Hymenomycotina</u>	
<b><u>Classe</u></b>	<u>Mammalia</u>	<u>Magnoliopsida</u>	<u>Homobasidiomycetae</u>	<u>Proteobacteria</u>
Subclasse	<u>Eutheria</u>	<u>Magnoliidae</u>	<u>Hymenomycetes</u>	<u>Gammaproteob</u>
<b><u>Ordem</u></b>	<u>Primatas</u>	<u>Fabales</u>	<u>Agaricales</u>	<u>Enterobacteriales</u>
Subordem	<u>Haplorrhini</u>	<u>Fabineae</u>	<u>Agaricineae</u>	
<b><u>Família</u></b>	<u>Hominidae</u>	<u>Fabaceae</u>	<u>Amanitaceae</u>	<u>Enterobacteriaceae</u>
Subfamília	<u>Homininae</u>	<u>Faboideae</u>	<u>Amanitoideae</u>	
<b><u>Género</u></b>	<u>Homo</u>	<u>Pisum</u>	<u>Amanita</u>	<u>Escherichia</u>
<b><u>Espécie</u></b>	<u>H. sapiens</u>	<u>P. sativum</u>	<u>A. muscaria</u>	<u>E. coli</u>

# Recentemente ...

O NCBI alcançou 1000 procariotos sequenciados <http://www.ncbi.nlm.nih.gov/>

Homens, animais, insetos e plantas pertencem ao domínio eucarionte <http://pt.wikipedia.org/wiki/Eucarionte>

## No ano de 1990 ...

Foi criado o Projeto Genoma Humano, com 3 milhões de dólares para realizar o seqüenciamento em 15 anos.

Foi criado um consórcio de várias universidades e institutos de pesquisa dos EUA e da Europa, com o objetivo de seqüenciar o DNA humano.

# O Projeto Genoma Humano

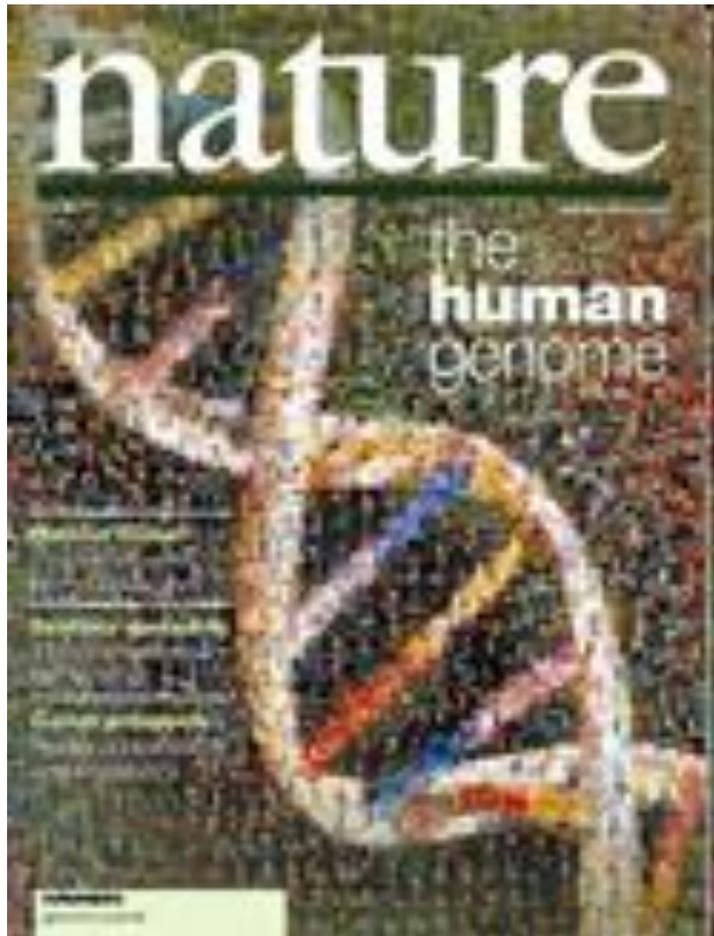
- O cronograma do projeto tinha previsão para conclusão em 2005.
- Um biólogo do NIH, Craig Venter, que havia seqüenciado a bactéria *Haemophilus influenzae* acreditou que era possível fazer o seqüenciamento em muito menos tempo, e resolveu criar a empresa *Celera Genomics*, com o objetivo de realizar isto, com menos tempo e 10% da verba.

# O Projeto Genoma Humano

- A proposta de Venter baseava-se na técnica de “*shot gun*”, já usada por ele no sequenciamento da influenza.
- Como estudo de caso, a Celera seqüenciou o genoma da *Drosophila melanogaster* (mosca) que tinha o comprimento de  $132 \times 10^6$  bases e aprox. 14 mil genes. O genoma da *Drosophila* foi publicado em 2000.
- Como fruto de um acordo, as seqüências do genoma humano da Celera e do Consórcio Mundial foram publicadas ao mesmo tempo, uma na revista *Science*, e a outra na revista *Nature*.

# O Genoma Humano é Publicado

**NATURE – 15 February 2001**  
**Volume 409 Number 6**



**SCIENCE - 16 February 2001**  
**Vol 291, Issue 5507**



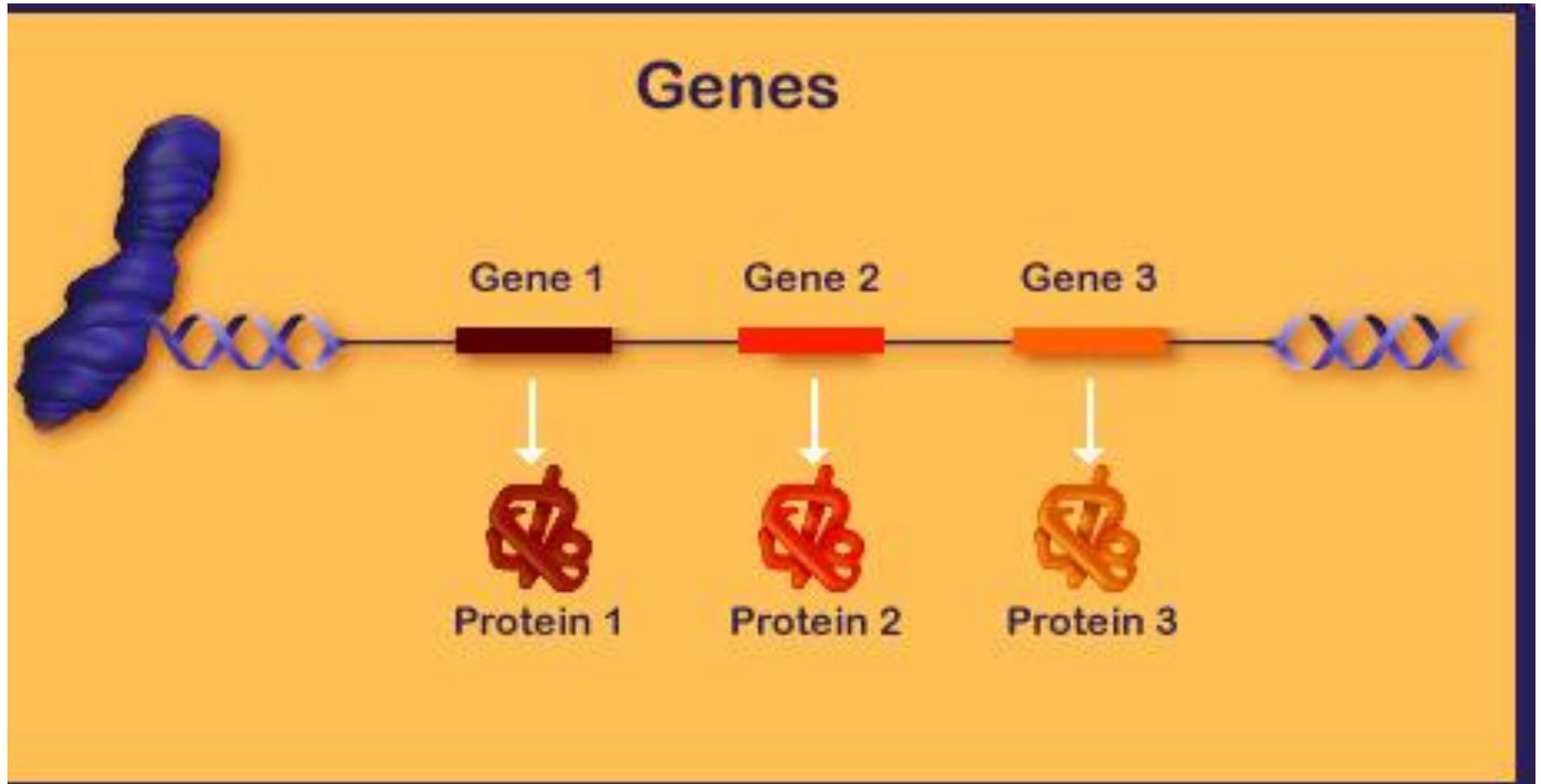
# Revelações do Genoma Humano

A seqüência do genoma humano tem cerca de  $3 \cdot 10^9$  nucleotídeos (caracteres em {A,C,G,T}), mas o número de genes esperado não foi encontrado.

→ Esperava-se algo da ordem de 80.000 a 100.000 mil genes

→ A anotação do seqüência do genoma humano tem cerca de 25.000 genes.

# Genes Codificam Proteínas



# Revelações do Genoma Humano

Como o número de proteínas conhecidas é muito maior, fica claro que algum tipo de controle faz com que um gene em geral se expresse como diferentes proteínas.

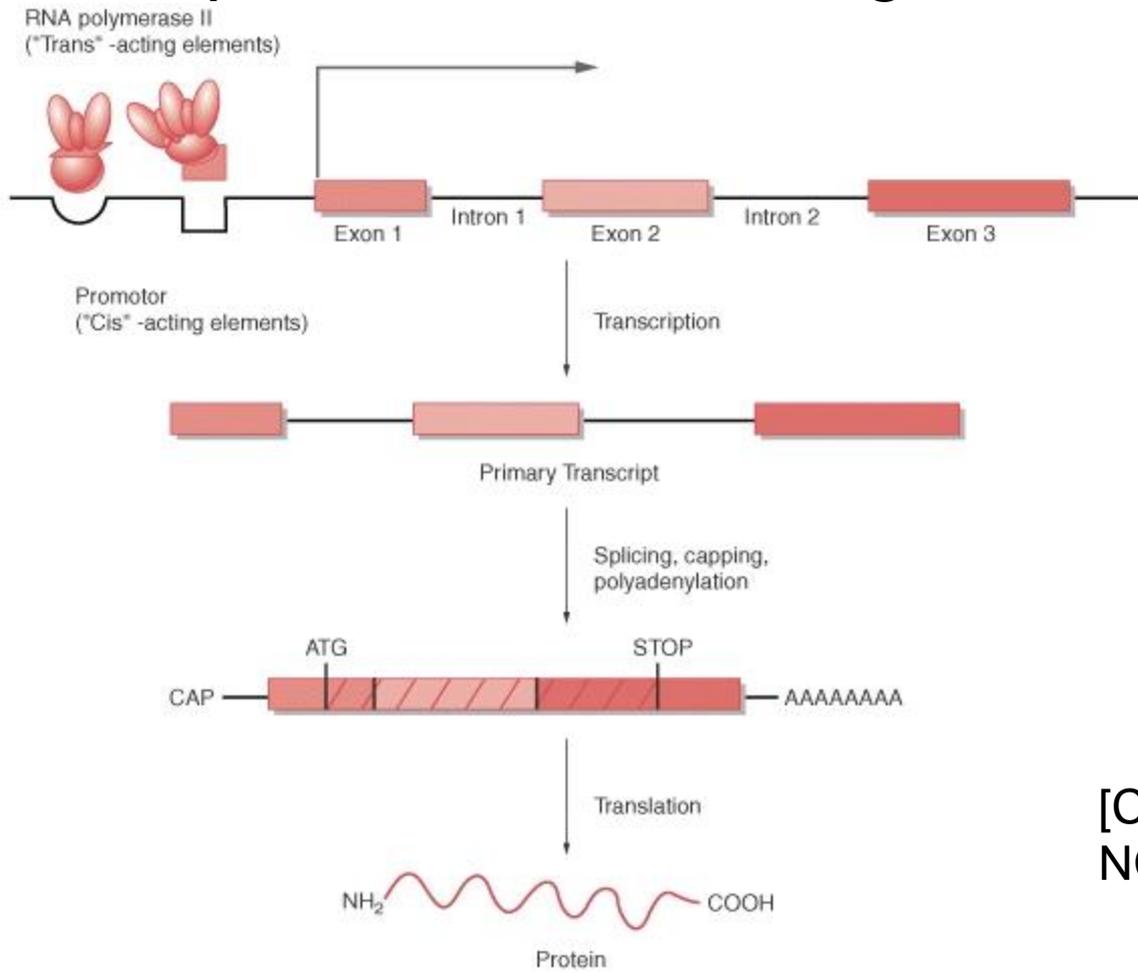
Uma seqüência de DNA:

[http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NT\\_009759.1](http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NT_009759.1)

# Como entender uma seqüência de DNA?

Procurando por sinais e características:

Ex: start codon, stop codon, splice donor e splice acceptor sites, TF binding sites, etc.



[Cancer Medicine]  
NCBI's Book Shelf

# Como capturar estes sinais?

Usando gene finding software:

<http://www.hku.hk/bruhk/sqgene.html>

Diagrama HMM

<http://genomebiology.com/2006/7/s1/S9/figure/F5>

# Como capturar estes sinais?

Uma alternativa:

Observando a conservação de trechos de aminoácidos

Primeira base ↓	2a. base				Terceira base ↓
	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

Alguns sinais não são tão característicos ...

E muitas vezes símbolos são trocados por sinônimos:

Para achá-los, é preciso alinhar seqüências e encontrar os chamados motifs:

<http://www.scielo.br/img/revistas/qn/v26n2/149982.gif>