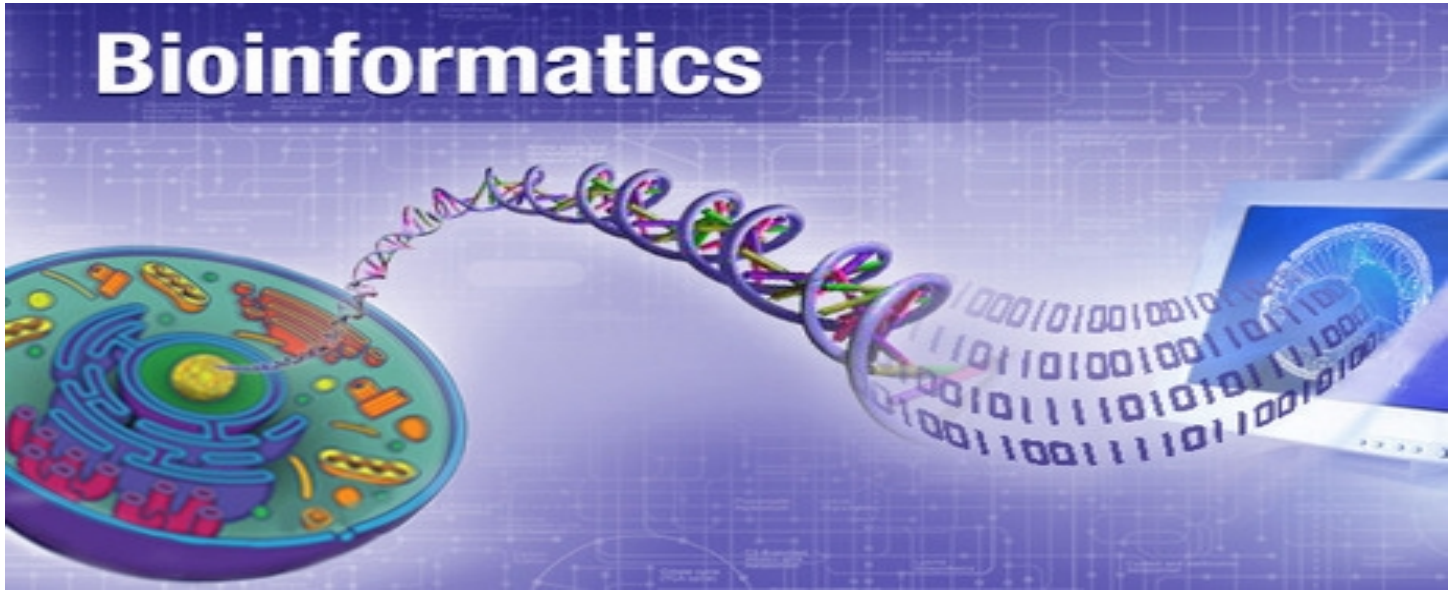
UFPR
UNIVERSIDADE FEDERAL DO PARANÁ



CI 242 – Tópicos de Pesquisa em Informática

Gerson Luiz dos Santos Junior

Lucas Alvarenga

Manuseio de Informações Biológicas

- O que é Bioinformática?
- Informática Clínica x Bioinformática
- Abordagem Reducionista x Modelos Integrativos

Fontes de Dados Biológicos

- 3 principais fontes:
  - Projeto Genoma Humano
  - Sequenciamento

  → Patógenos
  → Organismos modelo

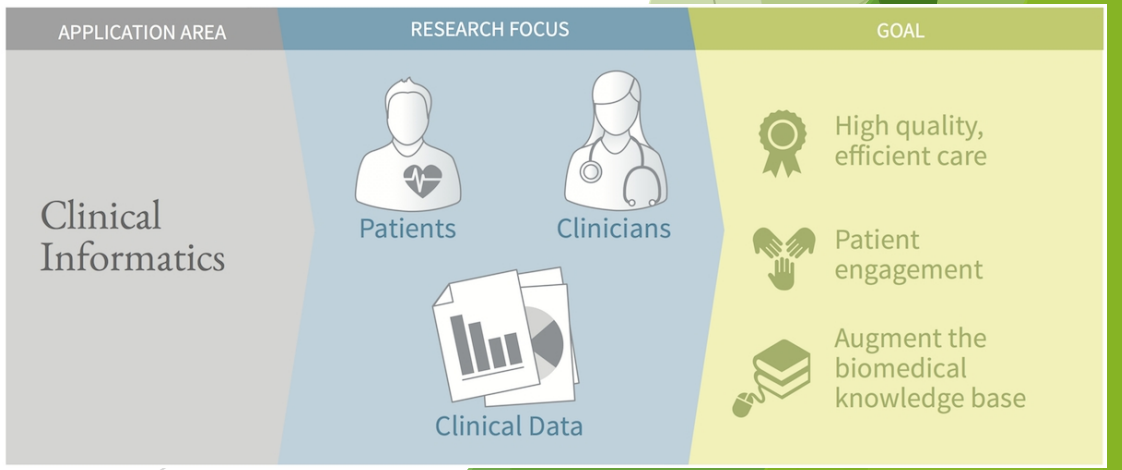
  } Genômica

  - Proteômica

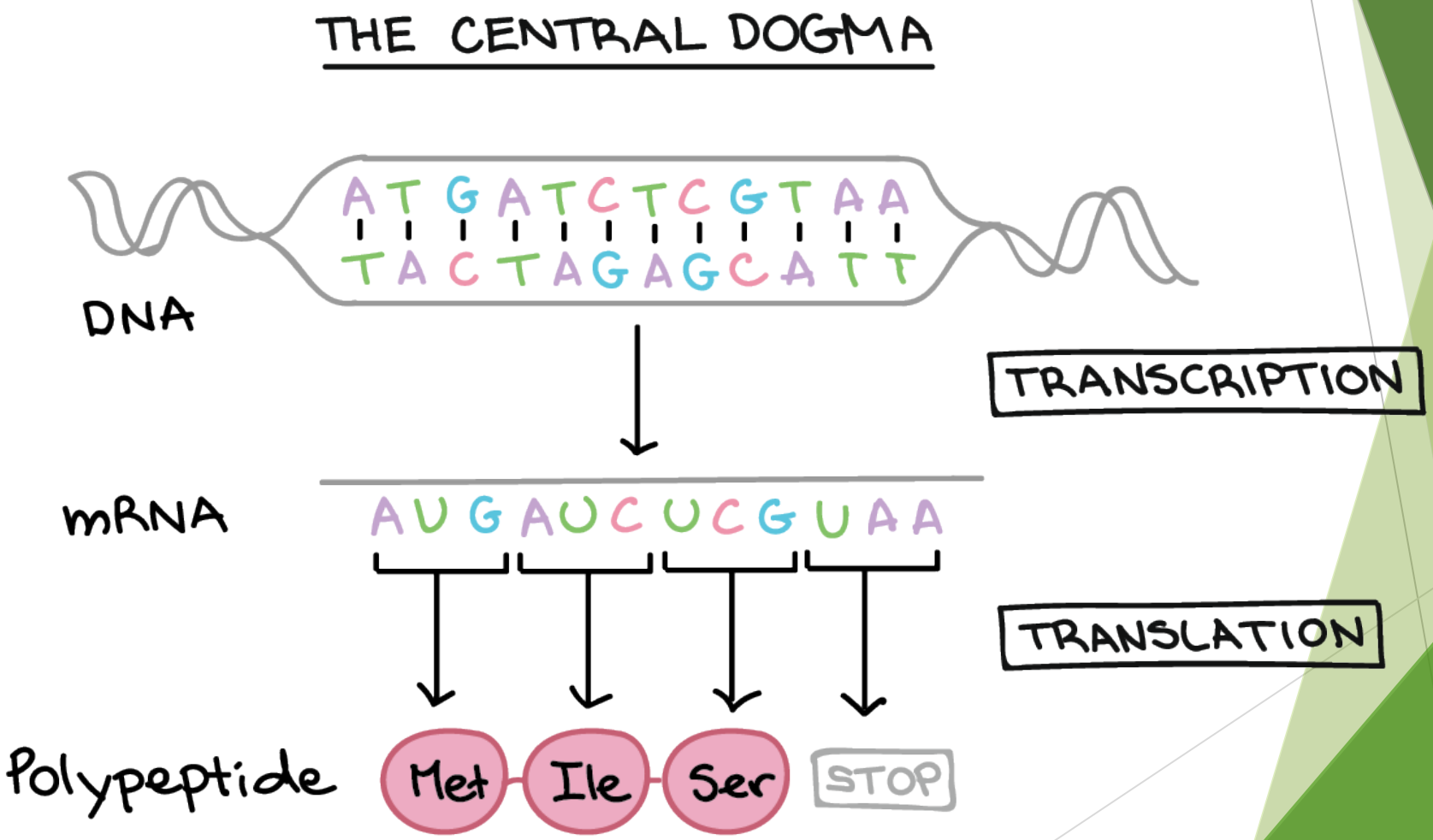
- MUITO MAIS FONTES!!
  - Surgimento das “Ômicas”

Implicações para a Informática Clínica

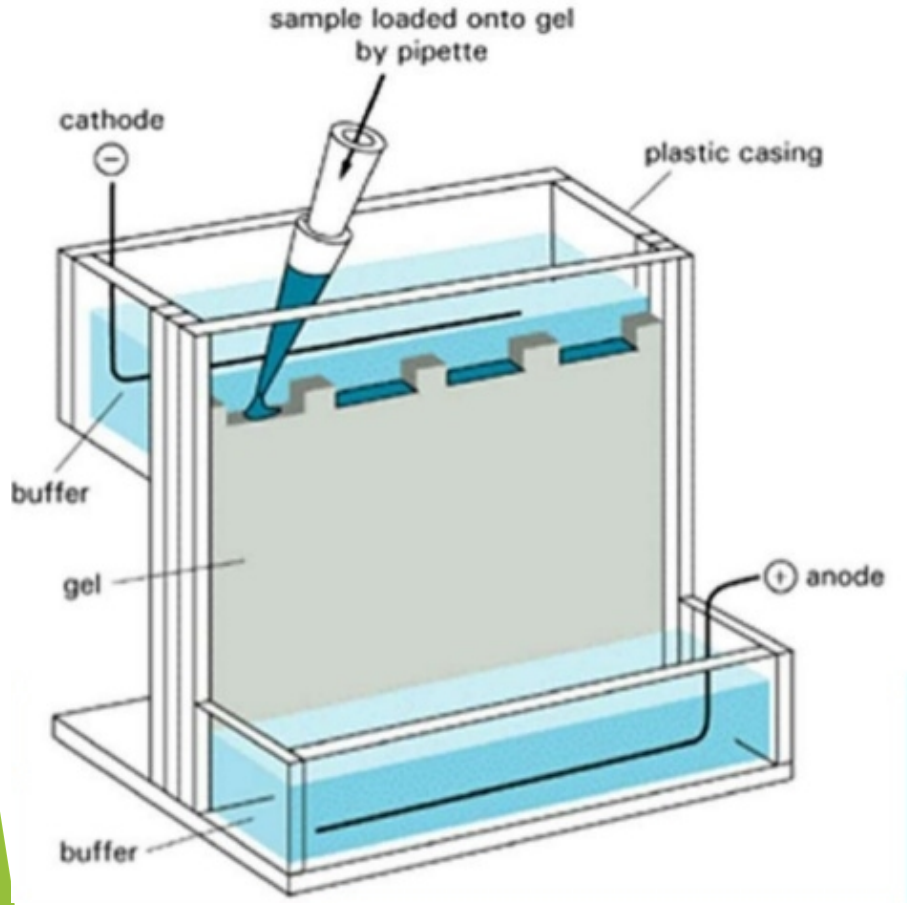
- Não é possível predizer precisamente!!
- Mudanças Possíveis:
  - 1 – Informações de sequência no prontuário do paciente;
  - 2 – Novas fontes de informações de diagnóstico e prognóstico;
  - 3 – Considerações Éticas.



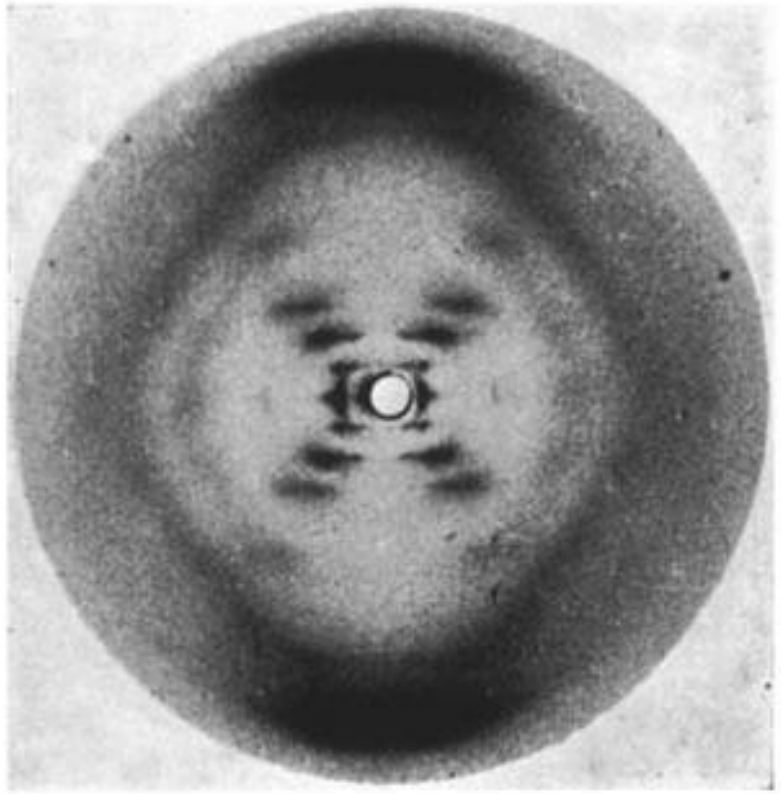
O Surgimento da Bioinformática



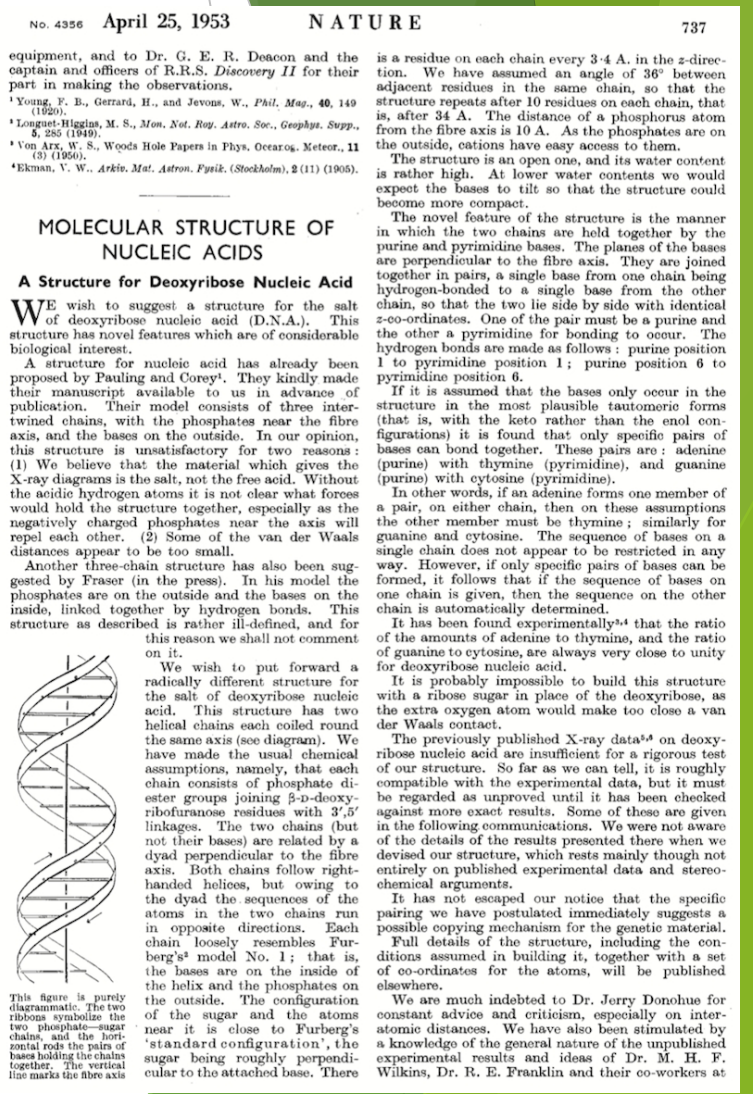
Raízes da Bioinformática Moderna



- Arne Tiselius



- Watson & Crick



No. 4356 April 25, 1953 NATURE 737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

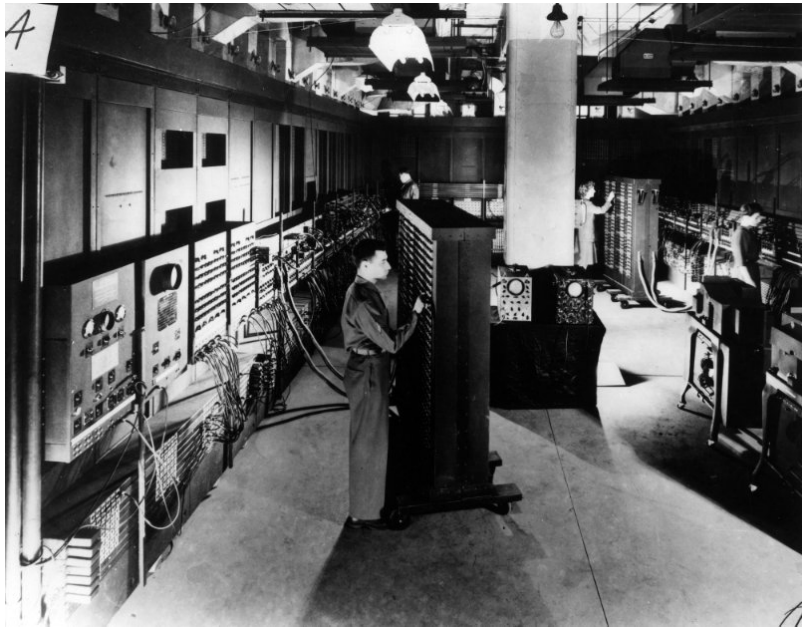
The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

Raízes da Bioinformática Moderna



Explosão de Dados

- Presença de computadores em diversos métodos:
  - Cristalografia de Raio-X, Ressonância Magnética Nuclear, Sequenciamento.

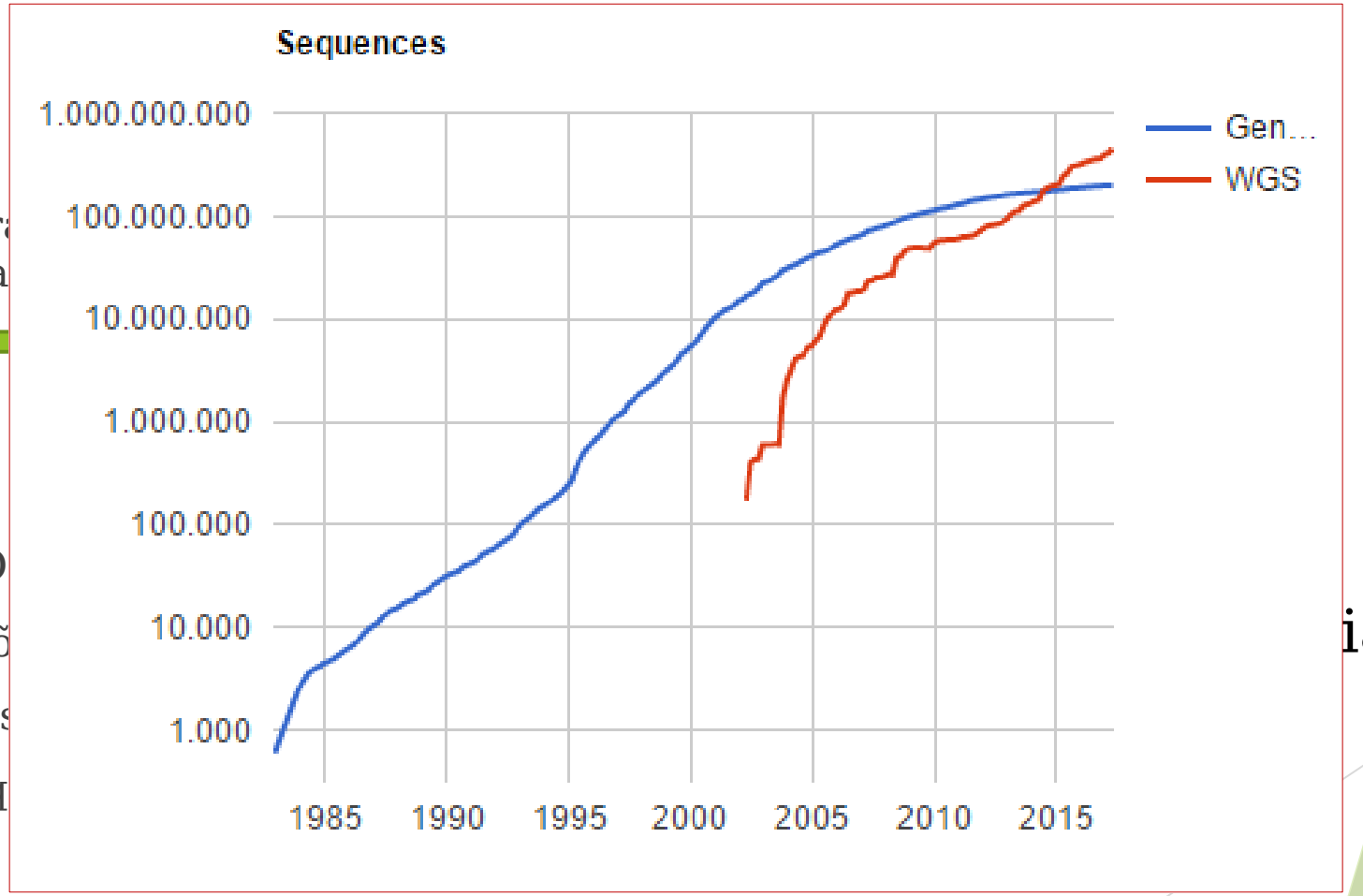
→ Armazenamento, Análise e Disseminação

- Dados

- Volume da Dados IMENSO!
  - 22.3 milhões de sequências → 451 milhões de sequências!!!
  - 15 milhões de citações literárias
  - 40 mil sequências proteicas → 130 mil sequências!!!

Explosão de Dados

- Presença de
  - Cristalogr
    Sequencia
- Dados
- Volume da D
  - 22.3 milhõ
  - 15 milhões
  - 40 mil seq



ias!!!

Explosão de Dados

- Presença de computadores em diversos métodos:
  - Cristalografia de Raio-X, Resonancia Magnética Nuclear, Sequenciamento.

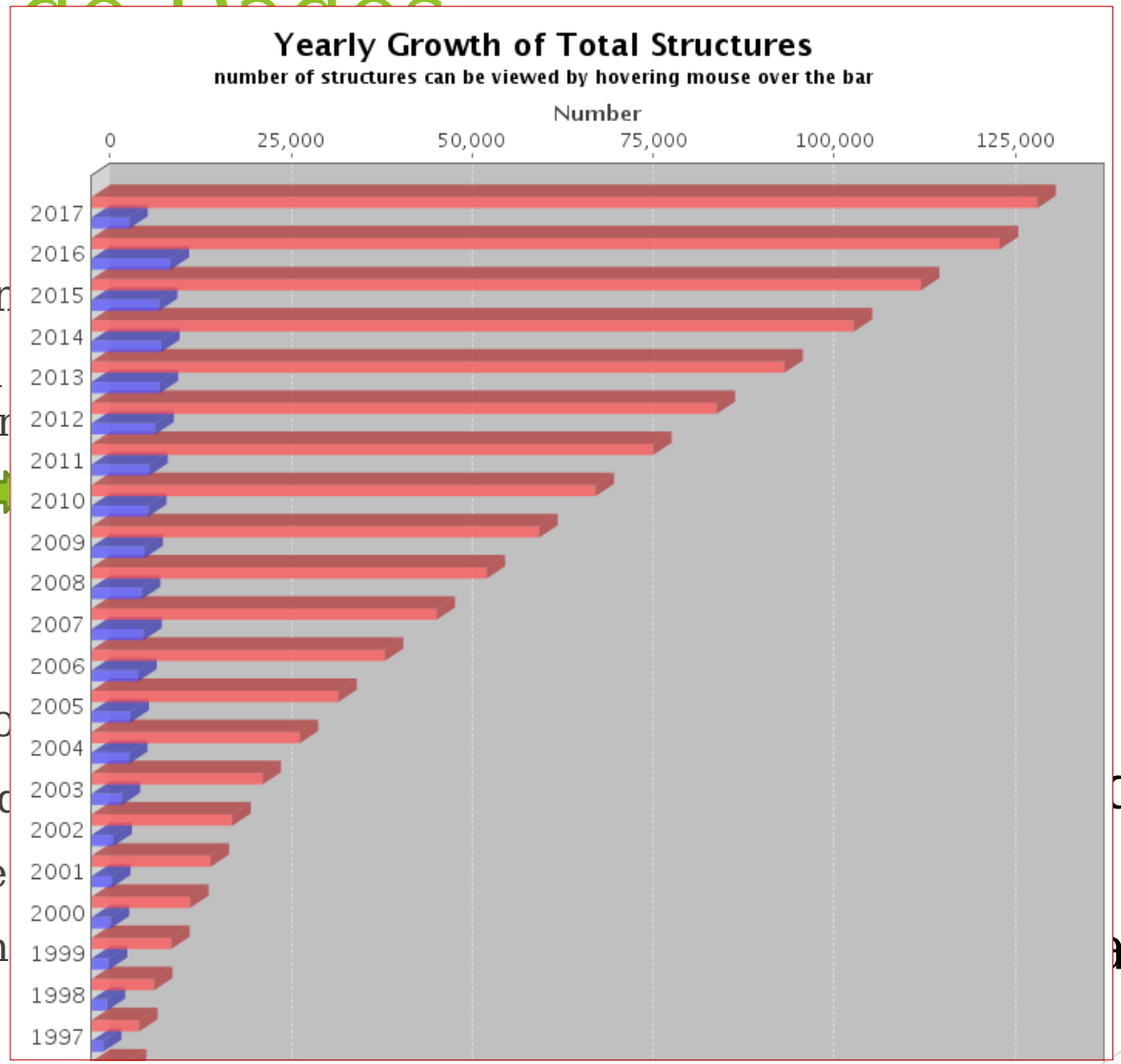
→ Armazenamento, Análise e Disseminação

- Dados

- Volume da Dados IMENSO!
  - 22.3 milhões de sequências → 451 milhões de sequências!!!
  - 15 milhões de citações literárias
  - 40 mil sequências proteicas → 130 mil sequências!!!

Explosão de Dados

- Presença de cor
  - Cristalografia
    Sequenciamer
- Dados
- Volume da Dado
  - 22.3 milhões d
  - 15 milhões de
  - 40 mil sequên



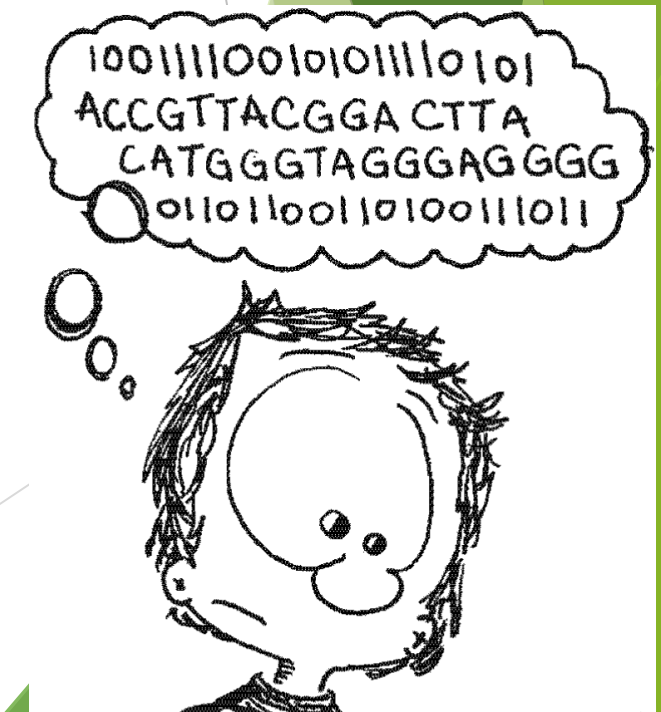
quências!!!

as!!!

Sequências na Biologia

- Blocos básicos de construção
- Problemas:
  - Modelos padrões de banco de dados;
    - Banco de Dados Relacional.
  - Sequência por si só
    - Banco de dados orientado a objetos;

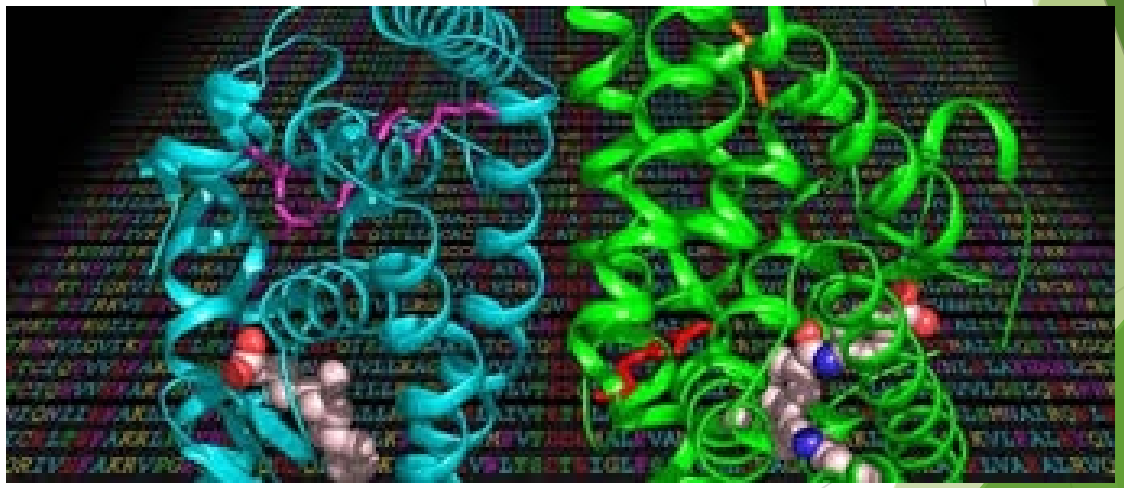
Sequência dentro de um grupo



Estruturas na Biologia

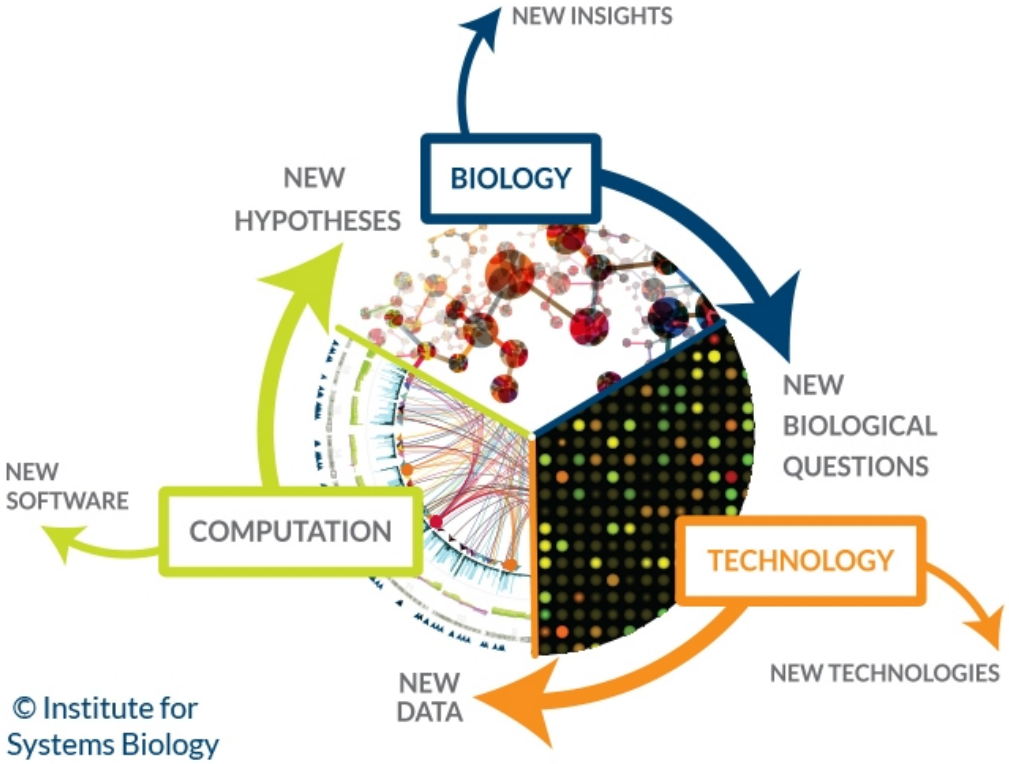
- Sequência 
- Estrutur 

- Estrutura leva à Função
- Densidade de informações!!



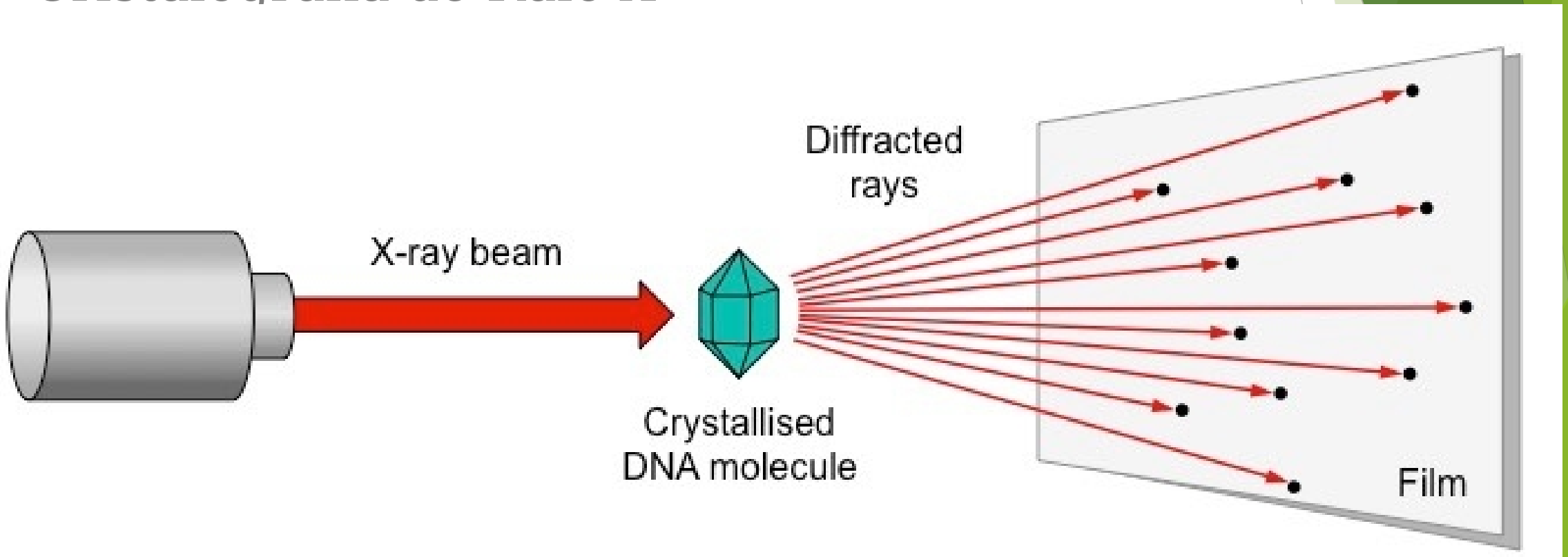
Biologia de Sistemas

- Entender como proteínas e genes interagem a nível celular.
- Algoritmos -> Análise integrada
  - Combate a doenças.
- Principais pesquisas na bioinformática
  - Vias chave de produção/degradação



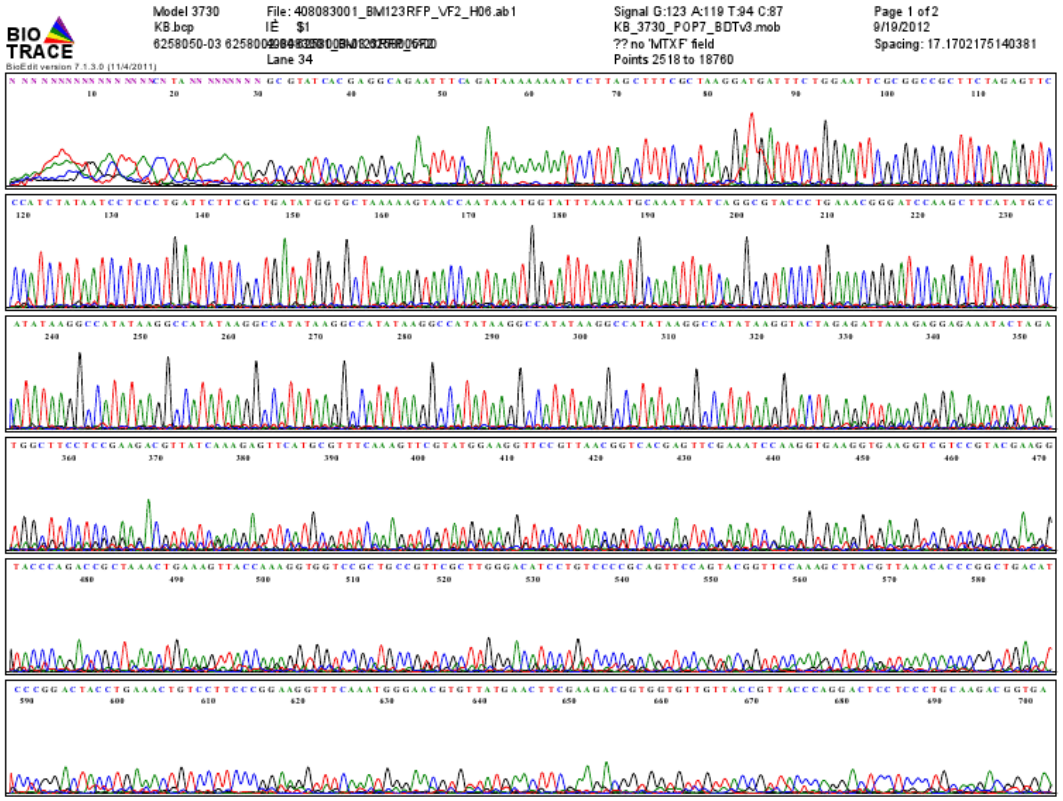
Análise Genômica

- Análise de estrutura por Cristalografia de Raio-X



Análise Genômica

- A Análise do Sequenciamento por BLAST



Download GenBank Graphics

Homo sapiens RNA, 45S pre-ribosomal 5 (RNA45S5), ribosomal RNA

Sequence ID: ref|NR_046235.1| Length: 13357 Number of Matches: 1

Range 1: 10714 to 10835 GenBank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
213 bits(115)	5e-52	120/122(98%)	1/122(0%)	Plus/Minus

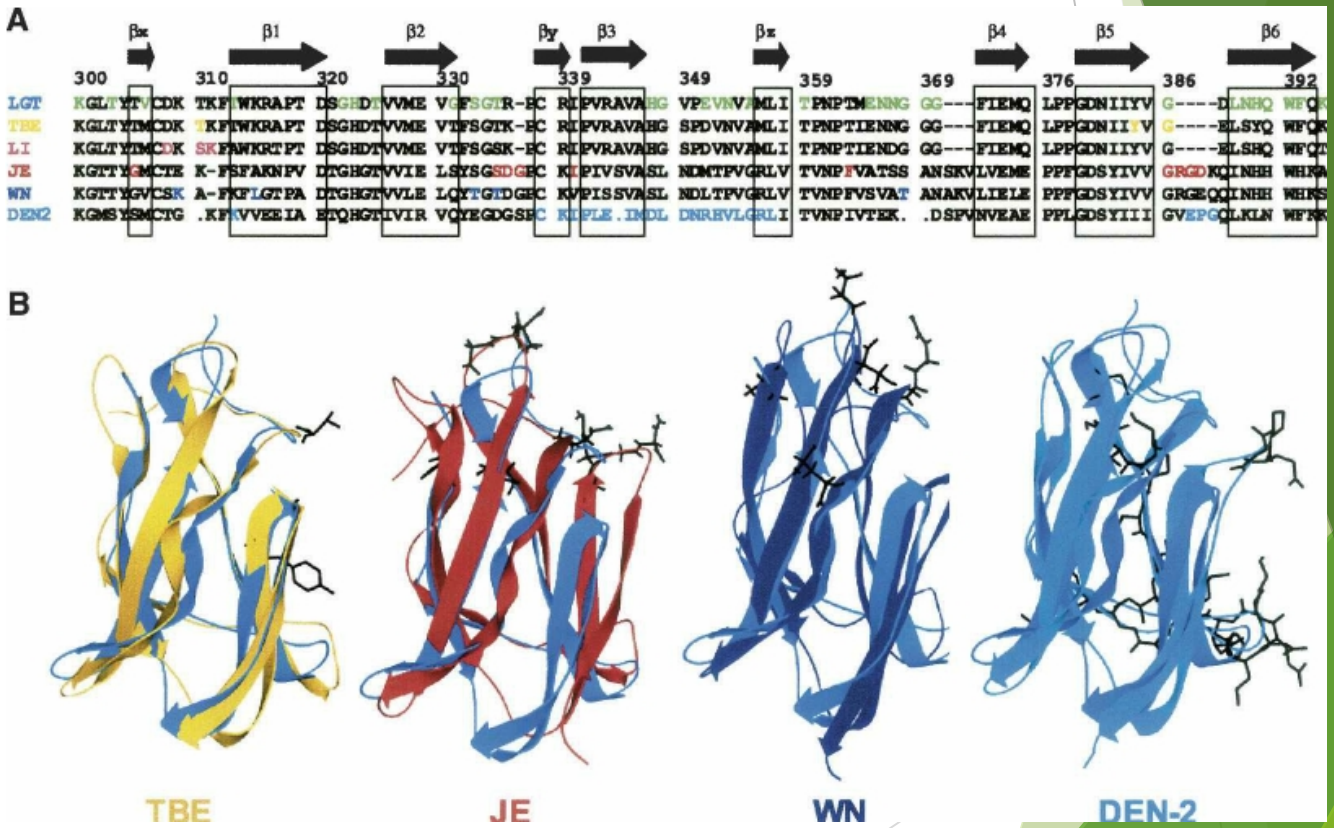
```
Query  39     CCGACCGACCCAGCCCTTAGAGCCAATCCTTATCCCGAAGTTACGGATCCGGCTTGCCGA  98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  10835  CCGACCGACCCAGCCCTTAGAGCCAATCCTTATCCCGAAGTTACGGATCCGGCTTGCCGA  10776

Query  99     CTTCCCTTACCTACATTGTTCCAACATGCCAGAGGCTGTTCACCTTGGAGACCTACTG-G  157
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||| | 
Sbjct  10775  CTTCCCTTACCTACATTGTTCCAACATGCCAGAGGCTGTTCACCTTGGAGACCTGCTGCG  10716

Query  158    GA  159
              ||
Sbjct  10715  GA  10714
```

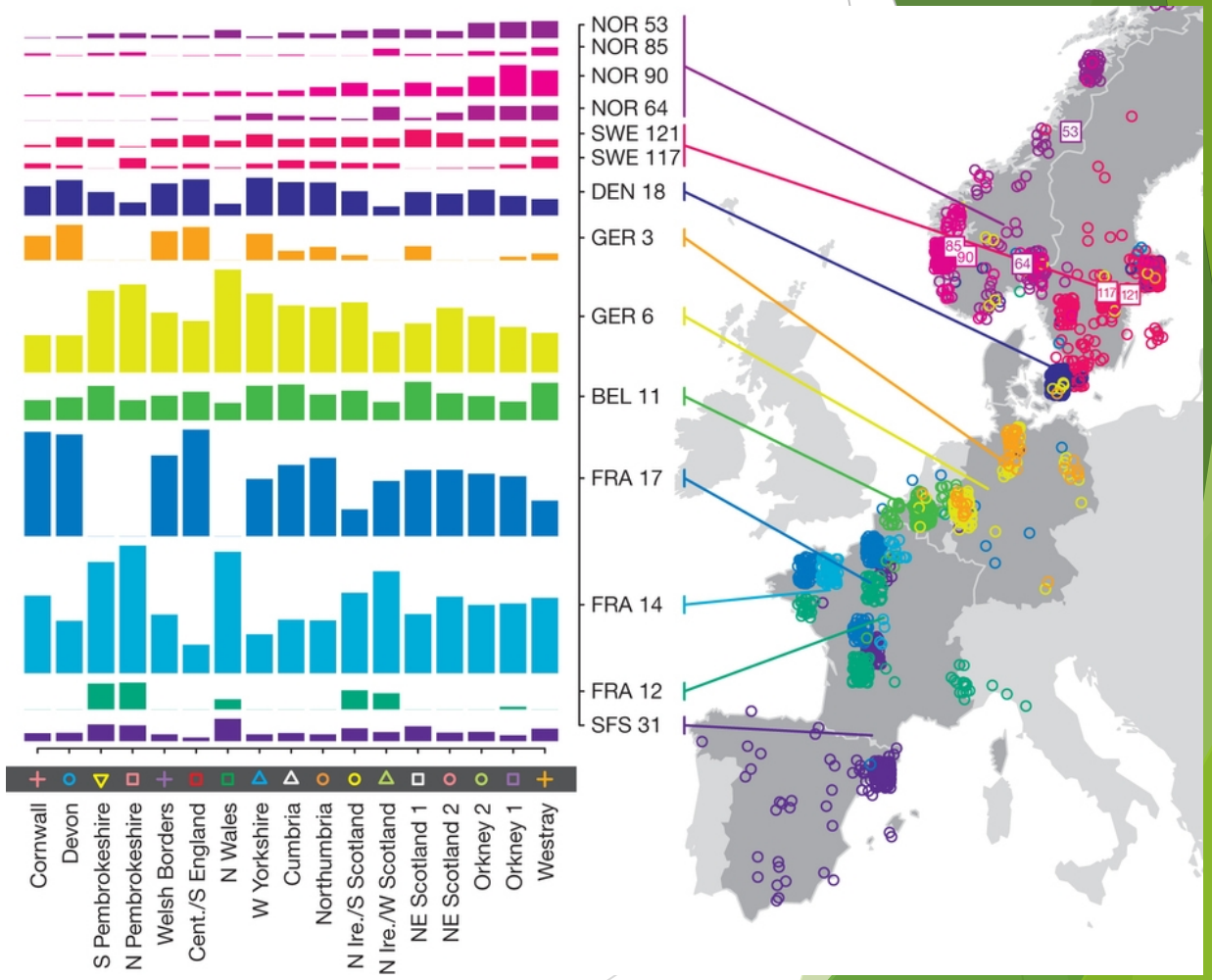
Análise Genômica

- A Predição da Estrutura e Função da Sequência



Análise Genômica

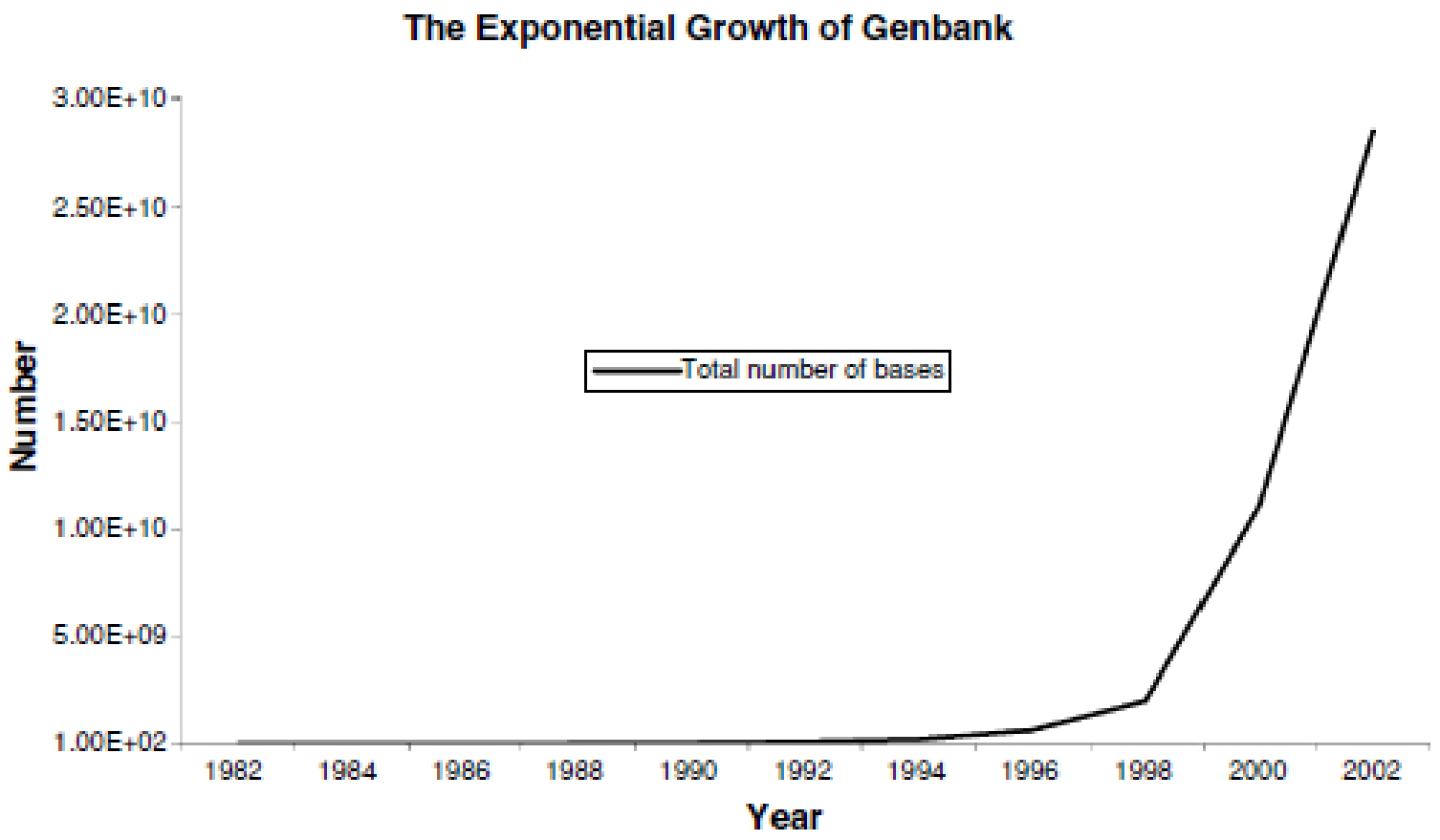
- Clustering da Expressão de Genes



Arquivo Genômica

- O GENBANK do NCBI

Arquivo Genômica



Arquivo Genômica

- ECOCYC:
Genoma E. coli

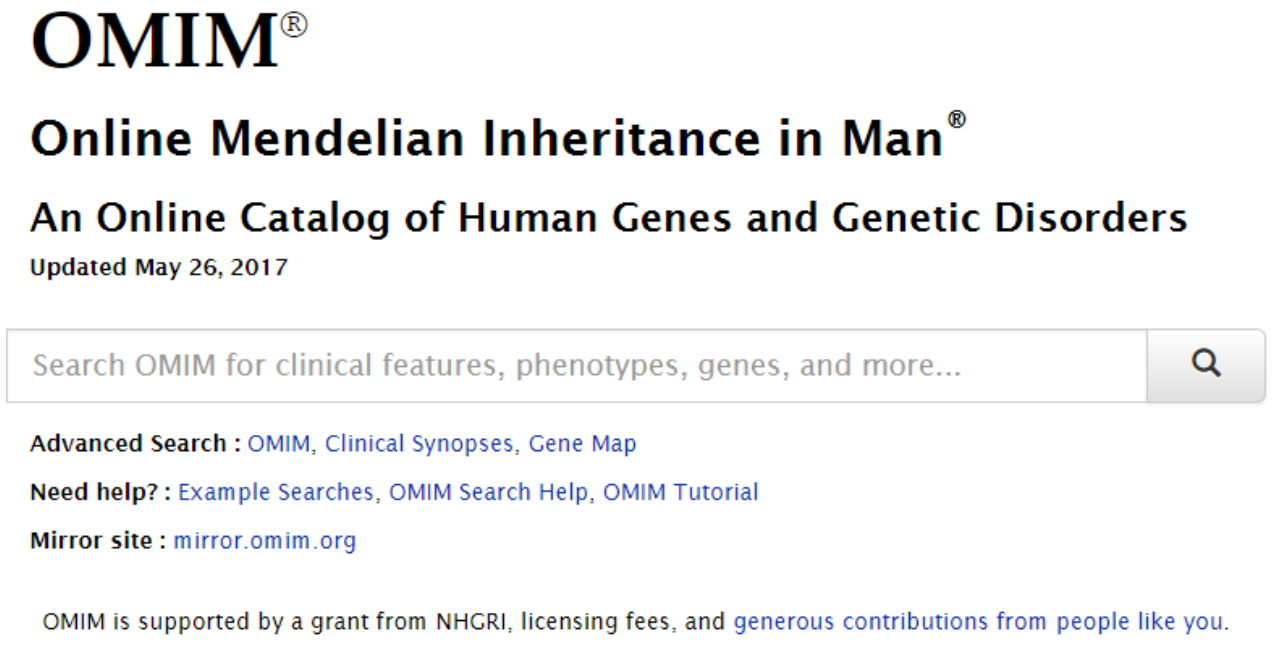
	Organism or Sample Properties
Relationship To Oxygen	facultative
Temperature Range	mesophile

Replicon	Total Genes	Protein Genes	RNA Genes	Pseudogenes	Size (bp)	NCBI Link
Chromosome	4494	4284	198	175	4,641,652	
Genes without a physical map position:	2					

Pathways:	343
Enzymatic Reactions:	1933
Transport Reactions:	477
Polypeptides:	4514
Protein Complexes:	1079
Enzymes:	1572
Transporters:	271
Compounds:	2751
Transcription Units:	3556
tRNAs:	89
Growth Media:	434
Transcriptional Regulation:	3553
Protein features:	4224
Phenotype Microarray Datasets:	5
GO Terms:	5739
Gene Essentiality Datasets:	5

Arquivo Genômica

- OMIM: Banco de Dados Pós-genômicos



Desafios

- Coletar o Genoma Humano de outros humanos, comparar e arquivar
- Associar a informação molecular com sintomas e doenças

Perguntas

1) Como a era da Genômica/Pós-Genômica ajudou a moldar as bases da Bioinformática? Comente um programa de Bioinformática que você já utilizou.

2) Graças à bioinformática, foram possíveis otimizar diversas etapas na área genômica. Descreva duas elas.