

**BIOINFORMÁTICA**

# INTRODUÇÃO

.O que é bioinformática?

A aplicação de técnicas computacionais visando o gerenciamento e a análise da informação oriunda na Biologia.

# OBJETIVO

- Criar uma representação visual do genoma e/ou código genético
- Investigar padrões
- Efetuar comparação entre representações.

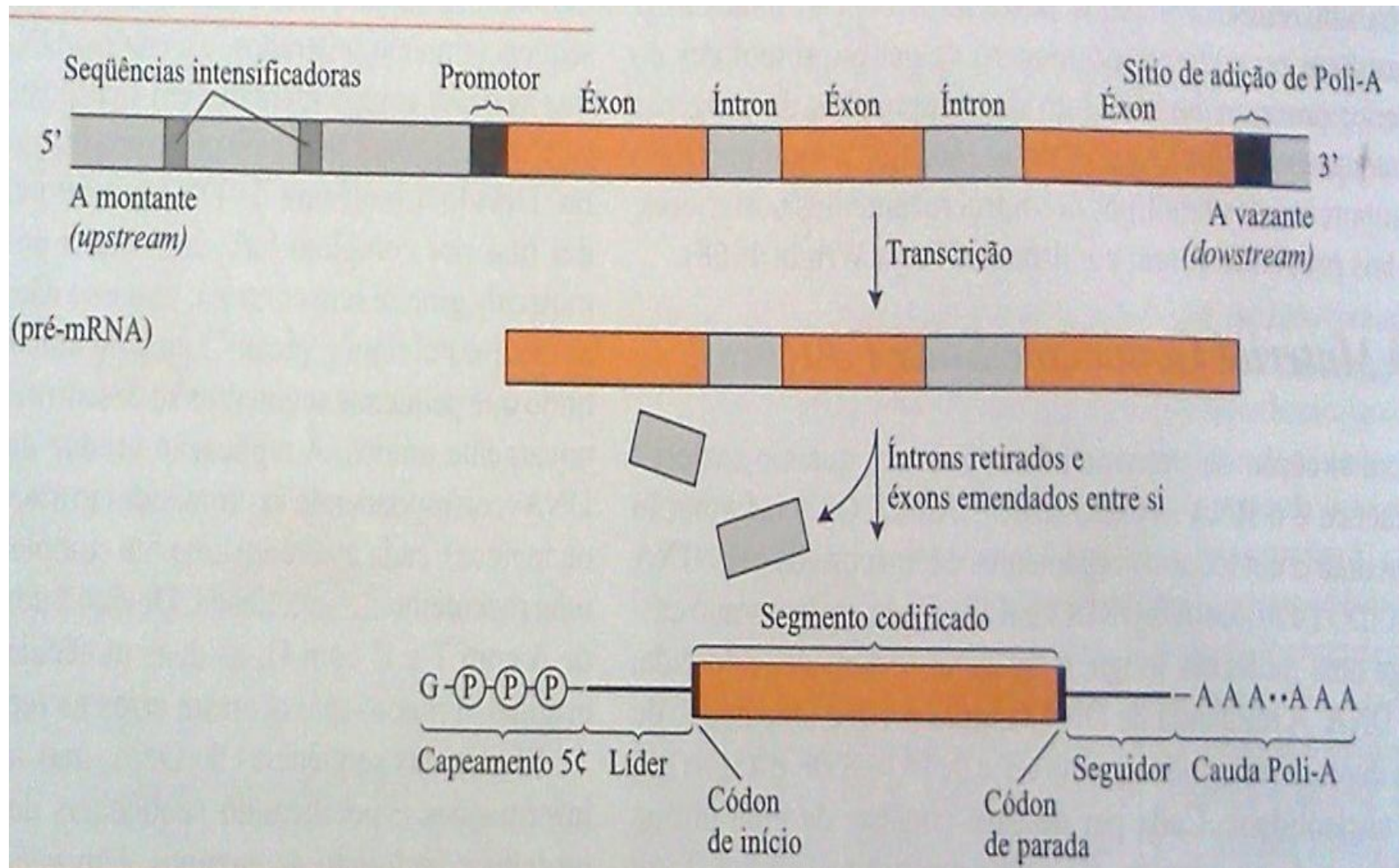
# JUSTIFICATIVA

## O APRENDIZADO DE GEOMETRIA É INTUITIVO



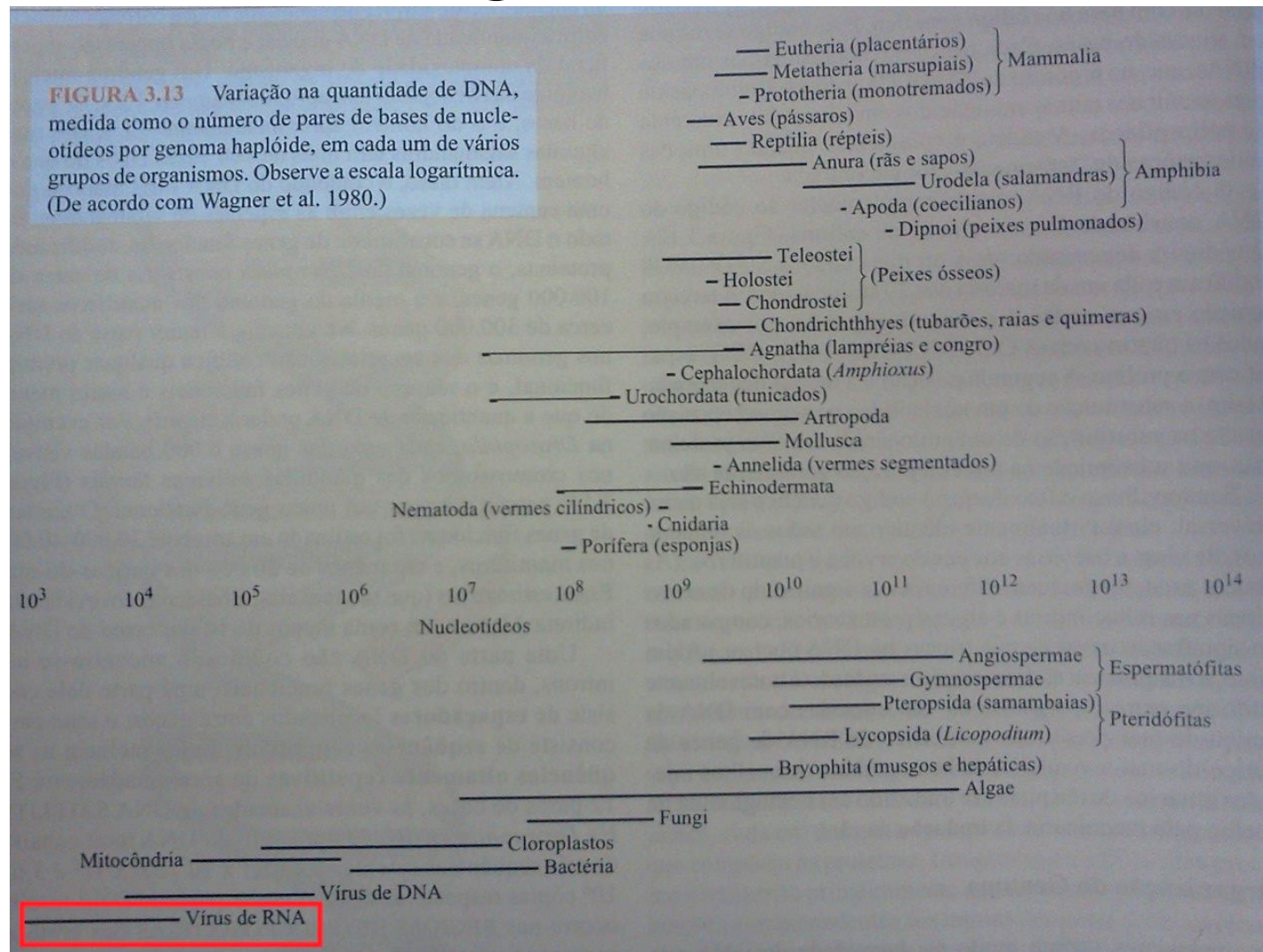
A comparação de formas é muito mais simplório do que a comparação de dados brutos.

# CÓDIGO GENÉTICO *versus* GENOMA





# Comparação dos tamanhos dos genomas



# COLETA DE DADOS

The screenshot shows the NCBI (National Center for Biotechnology Information) website. The browser's address bar at the top displays the URL [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov), which is circled in red. The website header includes the NCBI logo and a search bar. A left sidebar contains a 'Resource List (A-Z)' with links to various databases and tools. The main content area features a 'Welcome to NCBI' message, a 'Get Started' section with links to Tools, Downloads, How-To's, and Submissions, and a 'NCBI YouTube channel' advertisement. On the right, there are sections for 'Popular Resources' (listing PubMed, Bookshelf, etc.) and 'NCBI Announcements' (listing recent updates like the new Genome Workbench version).

NCBI Home  
Resource List (A-Z)  
All Resources  
Chemicals & Bioassays  
Data & Software  
DNA & RNA  
Domains & Structures  
Genes & Expression  
Genetics & Medicine  
Genomes & Maps  
Homology  
Literature  
Proteins  
Sequence Analysis  
Taxonomy  
Training & Tutorials  
Variation

**Welcome to NCBI**  
The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.  
[About the NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [RSS Feeds](#)

**Get Started**

- [Tools](#): Analyze data using NCBI software
- [Downloads](#): Get NCBI data or software
- [How-To's](#): Learn how to accomplish specific tasks at NCBI
- [Submissions](#): Submit data to GenBank or other NCBI databases

**NCBI YouTube channel**  
Learn how to get the most out of NCBI tools and databases with video tutorials on the NCBI YouTube Channel. [GO](#)

**Popular Resources**  
[PubMed](#)  
[Bookshelf](#)  
[PubMed Central](#)  
[PubMed Health](#)  
[BLAST](#)  
[Nucleotide](#)  
[Genome](#)  
[SNP](#)  
[Gene](#)  
[Protein](#)  
[PubChem](#)

**NCBI Announcements**  
New version of Genome Workbench available  
06 Sep 2012  
An integrated, downloadable application for reviewing and analyzing sequence data  
NCBI's July Newsletter is on the Bookshelf  
13 Aug 2012  
Introduction to the 1000 Genomes Browser, PubMed's Citation Manager and  
New Microbial BLAST Page  
28 Jun 2012  
Now easier to use and with the familiar

National Center for Biotechnology Information

www.ncbi.nlm.nih.gov

Genome

All Databases

PubMed

Protein

Nucleotide

GSS

EST

Structure

Assembly

BioProject

BioSample

BioSystems

Books

Conserved Domains

Clone

dbGaP

dbVar

Epigenomics

Gene

GEO DataSets

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

to NCBI

Center for Biotechnology Information advances science and health by providing access to biomedical information.

NCBI | Mission | Organization | Research | RSS Feeds

Popular Resources

PubMed

Bookshelf

PubMed Central

PubMed Health

BLAST

Nucleotide

Genome

SNP

Gene

Protein

PubChem

NCBI Facebook page

Find out the latest news about NCBI resources and participate in community discussions.

GO

1 2 3 4 5 6 7 8

NCBI Announcements

New version of Genome Workbench available

06 Sep 2012

An integrated, downloadable application for viewing and analyzing sequence data

NCBI's July Newsletter is on the Bookshelf

13 Aug 2012

Introduction to the 1000 Genomes Project, PubMed's Citation Manager and

New Microbial BLAST Page

28 Jun 2012

Now easier to use and with the familiar

18:53 28/11/2012



National Center for Biotechnology Information

Genome  Search

### NCBI Home

Resource List (A-Z)

- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

## Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [RSS Feeds](#)

### Get Started

- [Tools](#): Analyze data using NCBI software
- [Downloads](#): Get NCBI data or software
- [How-To's](#): Learn how to accomplish specific tasks at NCBI
- [Submissions](#): Submit data to GenBank or other NCBI databases

### NCBI Facebook page

Find out the latest news about NCBI resources and participate in community discussions.

[GO](#)

### Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

### NCBI Announcements

New version of Genome Workbench available  
06 Sep 2012

An integrated, downloadable application for viewing and analyzing sequence data

NCBI's July Newsletter is on the Bookshelf  
13 Aug 2012

Introduction to the 1000 Genomes Project, PubMed's Citation Manager and New Microbial BLAST Page  
28 Jun 2012

Now easier to use and with the familiar

Aguardando www.ncbi.nlm.nih.gov...

18:54 28/11/2012

hiv - Genome - NCBI

www.ncbi.nlm.nih.gov/genome/?term=hiv

NCBI Resources How To Sign in to NCBI

Genome Genome hiv Search

Save search Limits Advanced Help

Display Settings: Summary, 20 per page Send to: Filters: Manage Filters

Results: 1 to 20 of 54

<< First < Prev Page 1 of 3 Next > Last >>

**Human immunodeficiency virus 1**

1. **Human immunodeficiency virus 1** overview  
Kingdom: Viruses  
Chromosome: 1  
ID: 10319

**Simian-Human immunodeficiency virus**

2. **Simian-Human immunodeficiency virus** RefSeq Genome  
Kingdom: Viruses  
Chromosome: 1  
ID: 5174

**Human immunodeficiency virus 2**

3. **Human immunodeficiency virus 2** RefSeq Genome  
Kingdom: Viruses  
Chromosome: 1  
ID: 5133

**Mus musculus**

4. The laboratory mouse is a major model organism for basic mammalian biology, human disease, and genome evolution, and its genome has been sequenced  
Kingdom: Eukaryotes  
Haploid chromosomes: 21; Organelles: 1  
ID: 52

**Rattus norvegicus**

Find related data

Database: Select

Find items

Search details

"Human immunodeficiency virus 1"  
[Organism] OR "Human immunodeficiency virus 2"  
[Organism] OR "Simian-Human immunodeficiency virus"

Search See more...

Recent activity

Turn Off Clear

hiv (54)

Genome

See more...

Human immunodeficiency virus 1

www.ncbi.nlm.nih.gov/genome/10319

NCBI Resources How To Sign in to NCBI

Genome Genome Search Limits Advanced Help

Display Settings: Overview Send to:

Organism Overview; Genome Project Report; Genome Annotation Report

# Human immunodeficiency virus 1

Human immunodeficiency virus 1 overview

Lineage: Viruses[3201]; Retro-transcribing viruses[113]; Retroviridae[53]; Orthoretrovirinae[44]; Lentivirus[10]; Primate lentivirus group[4]; Human immunodeficiency virus 1[1]

Genome Sequencing Projects

Organism	BioProject	Status	Chrs	Size (Kb)	GC%	Gene	Protein	Nbrs
Human immunodeficiency virus 1	PRJNA15478	●	1	9.18	42.1	9	9	1,556
Human immunodeficiency virus 1	PRJNA41109	○	-	-	-	-	-	-

Genome Region

Go to nucleotide Graphics FASTA GenBank

Tools

- Genome Resource

Publications

- Signal peptide fragments of preprolactin and HIV-1 p-gp160 interact with calmodulin. Martoglio B, et al. EMBO J 1997 Nov 17

Related information

- BioProject
- Gene
- Other genomes for species
- Components
- Protein
- PubMed
- Taxonomy

Recent activity

Human immunodeficiency virus 1

hiv (54)

19:01 28/11/2012

Human immunodeficiency virus 1 complete genome 2 selected items - Nucleotide

www.ncbi.nlm.nih.gov/nuccore/9629357,4558520

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search

Limits Advanced Help

Display Settings: Summary, Sorted by Default order

Results: 2

Human immunodeficiency virus 1, complete genome

1. 9,181 bp linear RNA

Accession: NC\_001802.1 GI: 9629357

GenBank FASTA Graphics

HIV-1, complete genome

2. 9,181 bp linear RNA

Accession: AF033819.3 GI: 4558520

GenBank FASTA Graphics Related Sequences

Send to: Filter your results:

All (2)

Bacteria (0)

INSDC (GenBank) (1)

mRNA (0)

RefSeq (1)

Manage Filters

Analyze these sequences

Run BLAST

Find related data

Database: Select

Find items

Recent activity

Turn Off Clear

Human immunodeficiency virus 1

Genome

hiv (54)

Genome

See more...

www.ncbi.nlm.nih.gov/nuccore/NC\_001802.1

19:03 28/11/2012



Human immunodeficiency virus 1, complete genome

NCBI Reference Sequence: NC\_001802.1

GenBank Graphics

>gi|9629357|ref|NC\_001802.1| Human immunodeficiency virus 1, complete genome

GGTCTCTCTGGTTAGACAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCC  
TCAATAAAGCTTGCCTTGAGTGTCTCAAGTAGTGTGTGCCGCTCTGTTGTGTACTCTGGTAACTAGAGA  
TCCTCAGACCCCTTTAGTGTGTGGAAAAATCTCTAGCAGTGGCGCCCGAACAGGGACCTGAAAGCGAA  
AGGGAACACAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGG  
CGGCGACTGGTGTGACGCCAAAAATTTTACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCG  
TCAGTATTAAGCGGGGAGAAATTAGATCGATGGGAAAAATTCGGTTAAGGCCAGGGGAAAGAAAAAT  
ATAAATTAACACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTTCGAGTTAATCCTGGCCTGTTAGA  
AACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTT  
AGATCATTATATAACAGTAGCAACCTCTATTGTGTGCATCAAAGGATAGAGATAAAGACACCAAGG  
AAGCTTTAGACAGATAGAGGAAGAGCAAAACAAAGTAAGAAAAAGCACAGCAAGCAGCAGCTGACAC  
AGGACACAGCAATCAGGTTCAGCCAAAATTACCTTATAGTGCAGAACATCCAGGGGCAATGGTACATCAG  
GCCATATCACCTAGAACTTTAAATGCATGGGTAAGTAGTAGAAGAGAGGCTTTCCAGCCAGAAAGTGA  
TACCCATGTTTTCAGCATTATCAGAAGGAGCCACCCACAAAGATTAAACACCATGCTAAACACAGTGGG  
GGGACATCAAGCAGCCATGCAATGTTAAAGAGACCATCAATGAGGAAGCTGCAGATGGGATAGAGTG  
CATCCAGTGCATGCAAGGCTATTGCACCAAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGAA  
CTACTAGTACCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCTATCCAGTAGGAGAAATTTA  
TAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTACCAAGCATTCTGGACATA  
AGACAAGGACCAAGGAACCTTTAGAGACTATGTAGACCGGTTCTATAAACTCTAAGAGCCGAGCAAG  
CTTCACAGGAGGTAAGAAATTTGGATGACAGAAACCTTGTGTGTCAAAATGCGAACCCAGATTGTAGAC  
TATTTTAAAGCATTGGGACAGCGGCTACACTAGAAGAAATGATGACAGCATGTCCAGGAGTAGGAGGA  
CCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATTCAGCTACCATATATGATGC  
AGAGAGGCAATTTAGGAACCAAGAAAGATTGTTAAGTGTTCATTGTGGCAAGAAAGGGCACACAGC  
CAGAAATTCAGGGGCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAATGAAAGAT  
TGTAAGTGTAGAGACAGGCTAATTTTTAGGGAAGATCTGGCCTTCCTACAAGGAAGGCCAGGGAATTTTC  
TTCAGAGCAGACAGGACCAACAGCCCCACCAGAGAGAGCTTCAGGTCTGGGGTAGAGACAACTCC  
CCCTCAGAGCAGGAGCCGATAGACAAGGAAGTATCTTAACTCCCTCAGGTCACTCTTTGGCAAC

APROXIMADAMENTE  
10 MIL BASES  
NITROGENADAS

Display Settings: FASTA

Send:

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Retrovirus Resource

Information and tools to support retrovirus research.

LinkOut to external resources

LANL HIV Databases [LANL HIV Databases]

MuLDS HIV-1 nucleotide subtype prediction [Multidimensional scaling and ...]

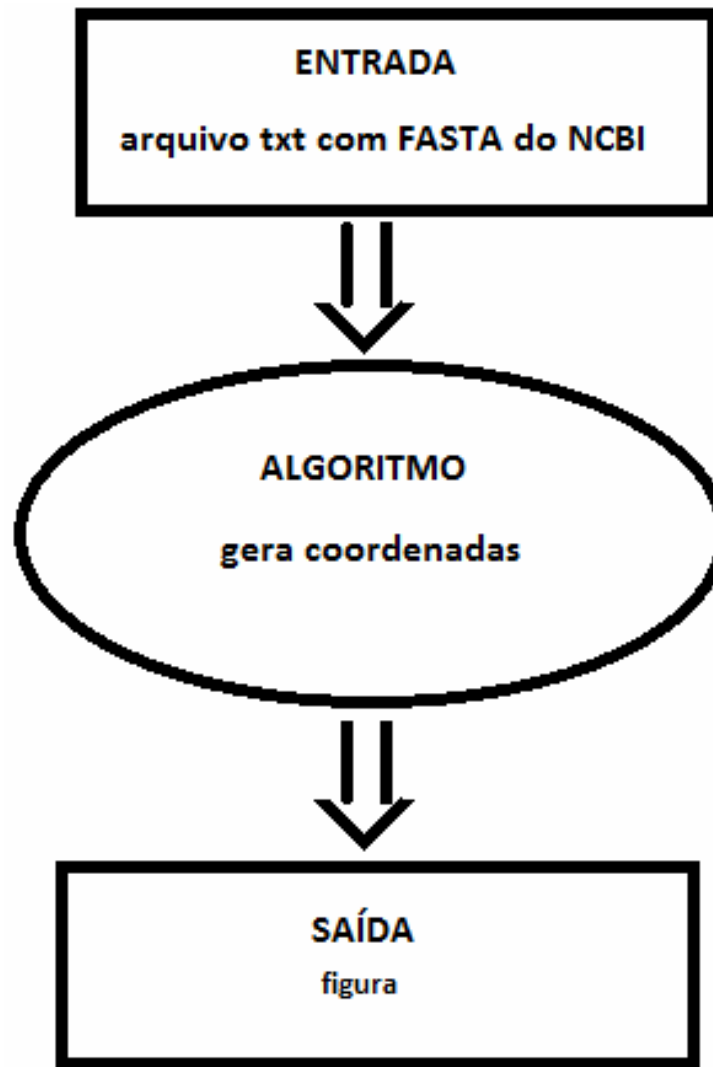
Related information

BioProject

Full text in PMC

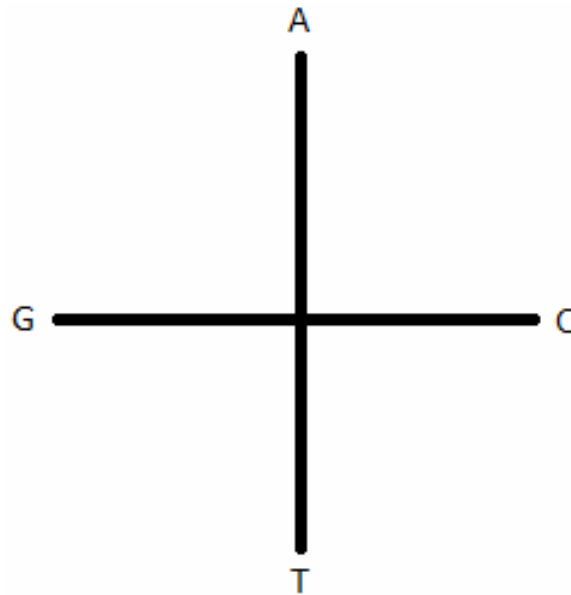
19:04  
28/11/2012

# FLUXOGRAMA



# ALGORITMO

Inspirado no RoboMind, conforme “rosa dos ventos” a seguir:



# Código Scilab

- A princípio, sequencia de letras foi substituída pelos seguintes números:

*// para a base A --> i*

*// para a base T --> -i*

*// para a base C --> 1*

*// para a base G --> -1*

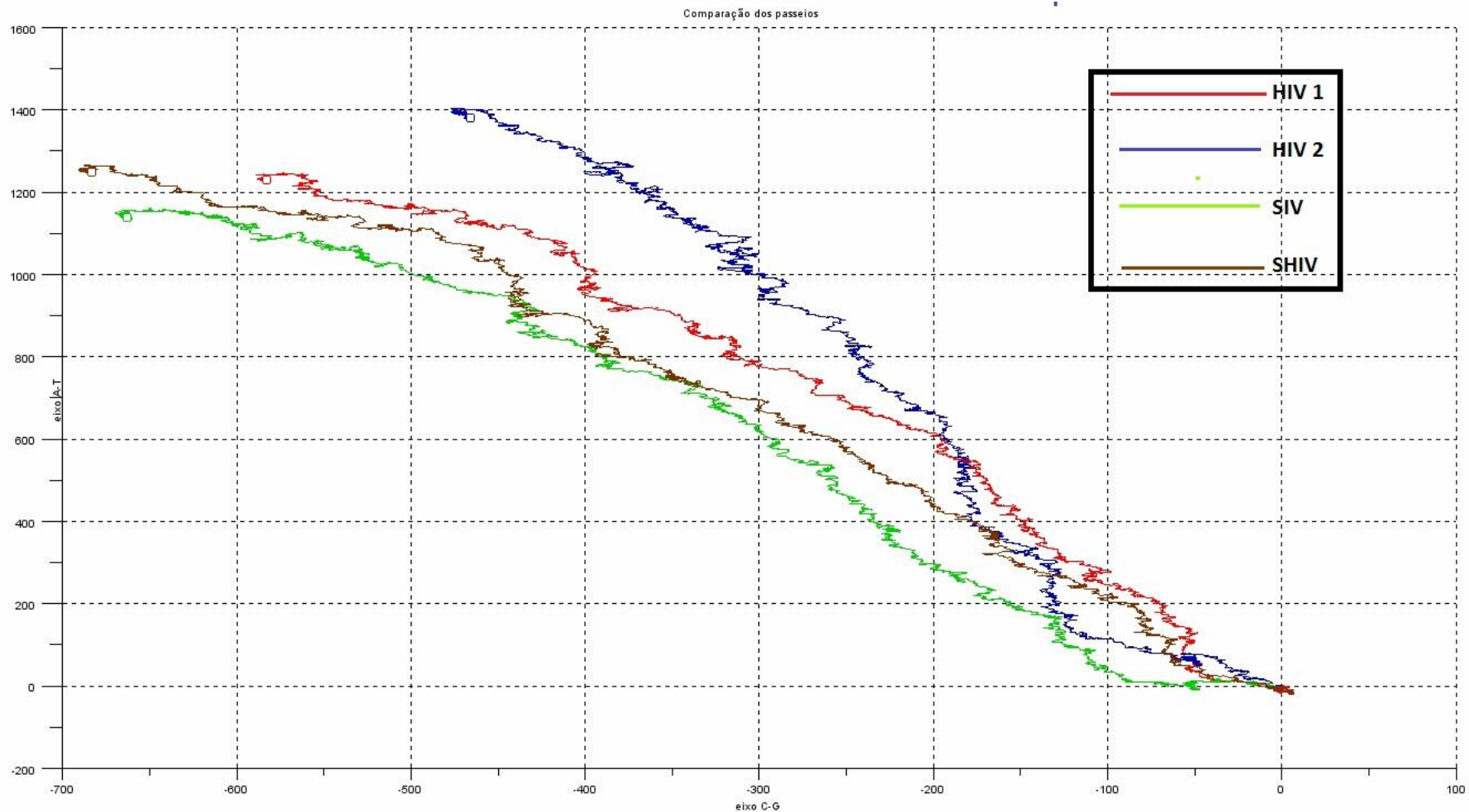
*// para base não reconhecida --> 0*



- Estes números foram interpretados como deslocamentos no plano complexo.

```
if part(genoma, k) == 'A'
    Z(k+1) = Z(k) + %i; // anda "Norte"
else
    if part(genoma, k) == 'T'
        Z(k+1) = Z(k) - %i; // anda "Sul"
    else
        if part(genoma, k) == 'C' then
            Z(k+1) = Z(k) + 1; // anda "Leste"
        else
            if part(genoma, k) == 'G' then
                Z(k+1) = Z(k) - 1; // anda "Oeste"
            else
                Z(k+1) = Z(k); // não anda (caso genoma(k) == 'N')
            end
        end
    end
end
end
end
```

# RESULTADOS



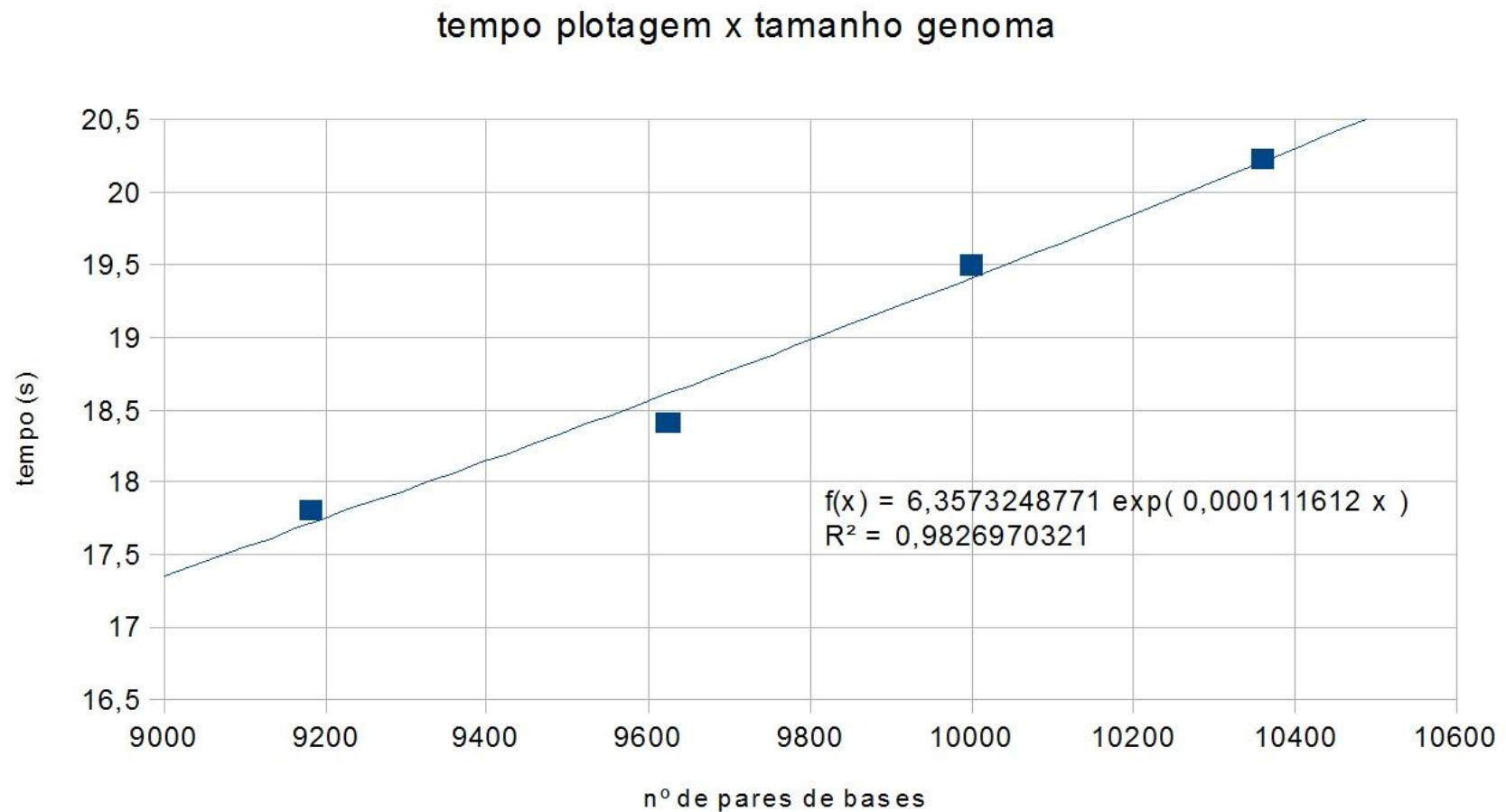
# CONCLUSÃO

- O caminho vermelho é mais próximo do verde-limão

Ou seja

- O vírus tipo 1 tem um genoma mais parecido com o tipo símio (SIV) do que o tipo 2.

# Métrica de desempenho: tempo





# Extrapolação para o genoma humano

- No genoma humano há aproximadamente 3000000000 de pares de bases nitrogenadas, assim, podemos estimar em quanto tempo o Scilab demoraria para plotar seu gráfico.

$$6,3573248771 \cdot e^{334836} \text{ s}$$

# Referências

- FUTUYMA, Douglas J.. **Biologia Evolutiva**. Tradução: Iulo Feliciano Afonso. 3 ed. Ribeirão Preto: 2009. 830p.
- [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)
- WATSON, BAKER, BELL, et al. **Biologia molecular do gene**. Tradução: Luciane Passaglia. 5 ed. São Paulo: 2006. 728p.