

# **Análise bioinformática de genomas de parasitas e organismos modelo**

**“Ferramentas e abordagens computacionais  
para estudos funcionais e análises da  
estrutura proteica”**

Processos  
biológicos  
específicos

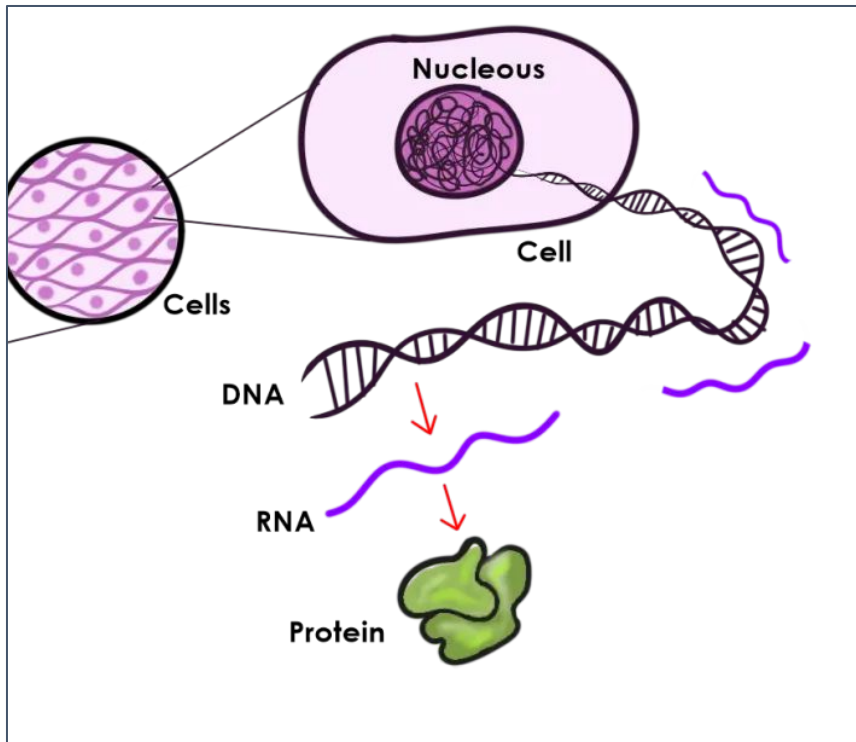
Manutenção  
Reprodução

Similaridade de  
características  
genéticas

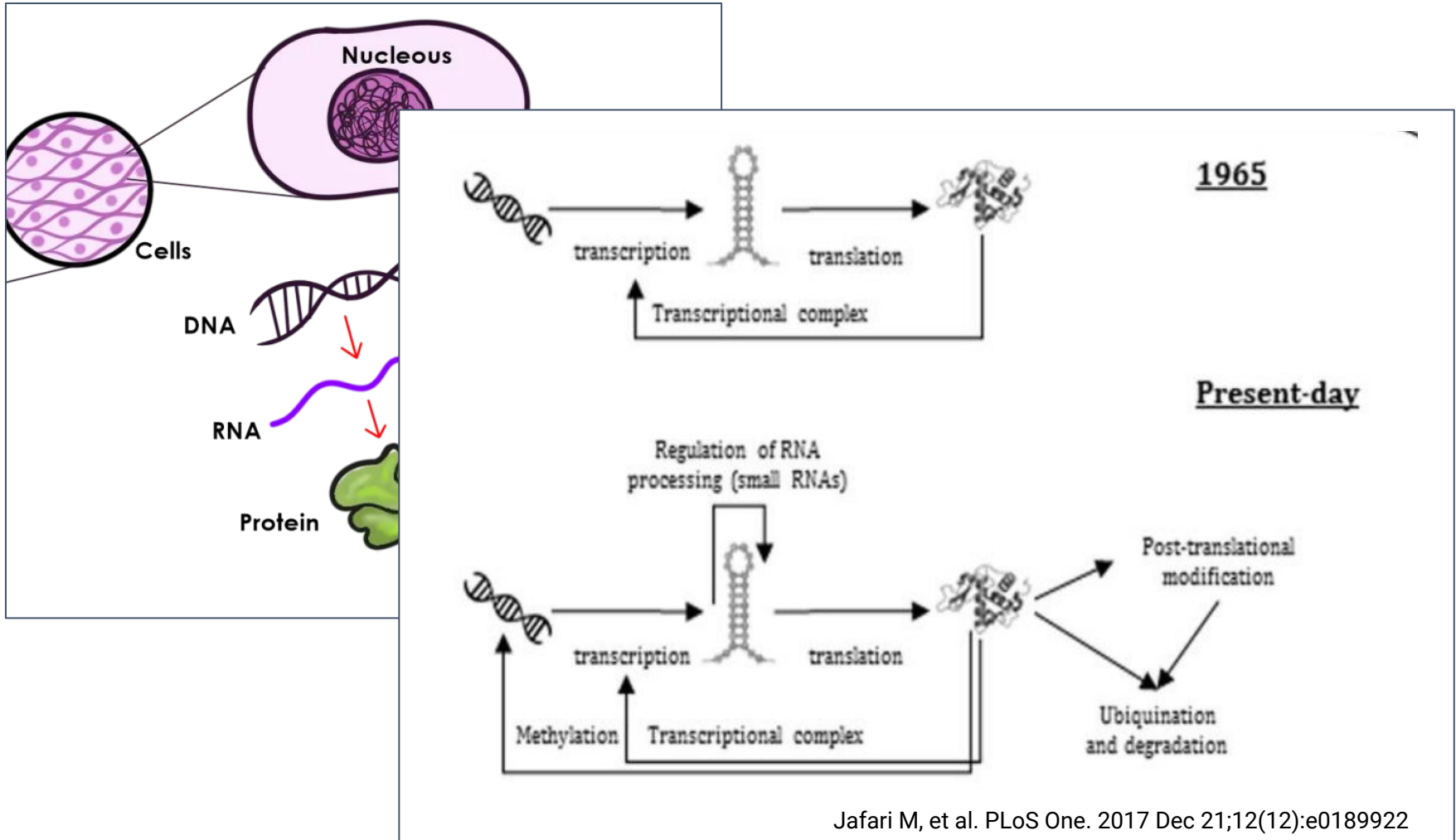
## Organismos modelo na pesquisa



## Dogma Central da Biologia Molecular

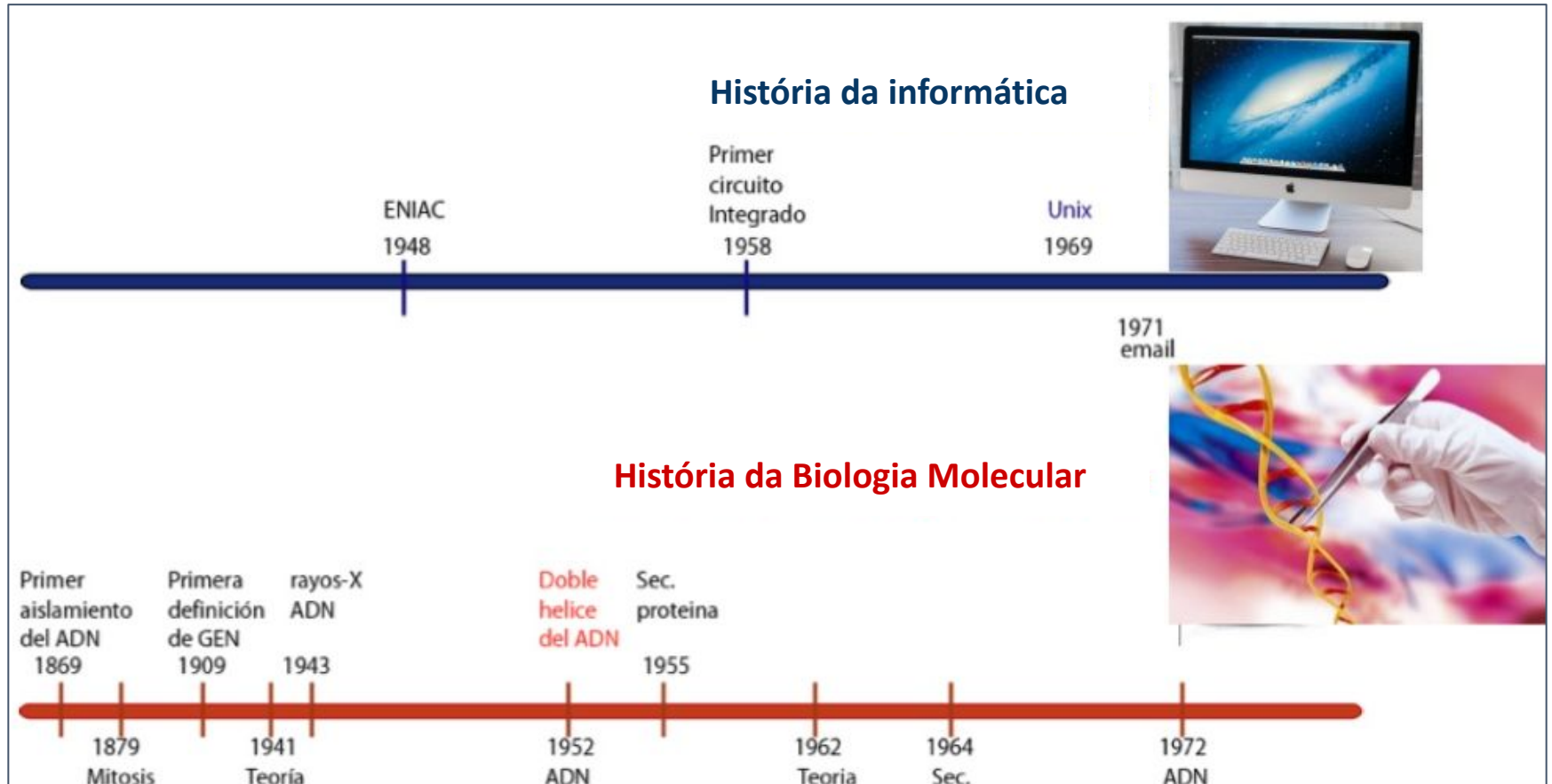


## Dogma Central da Biologia Molecular

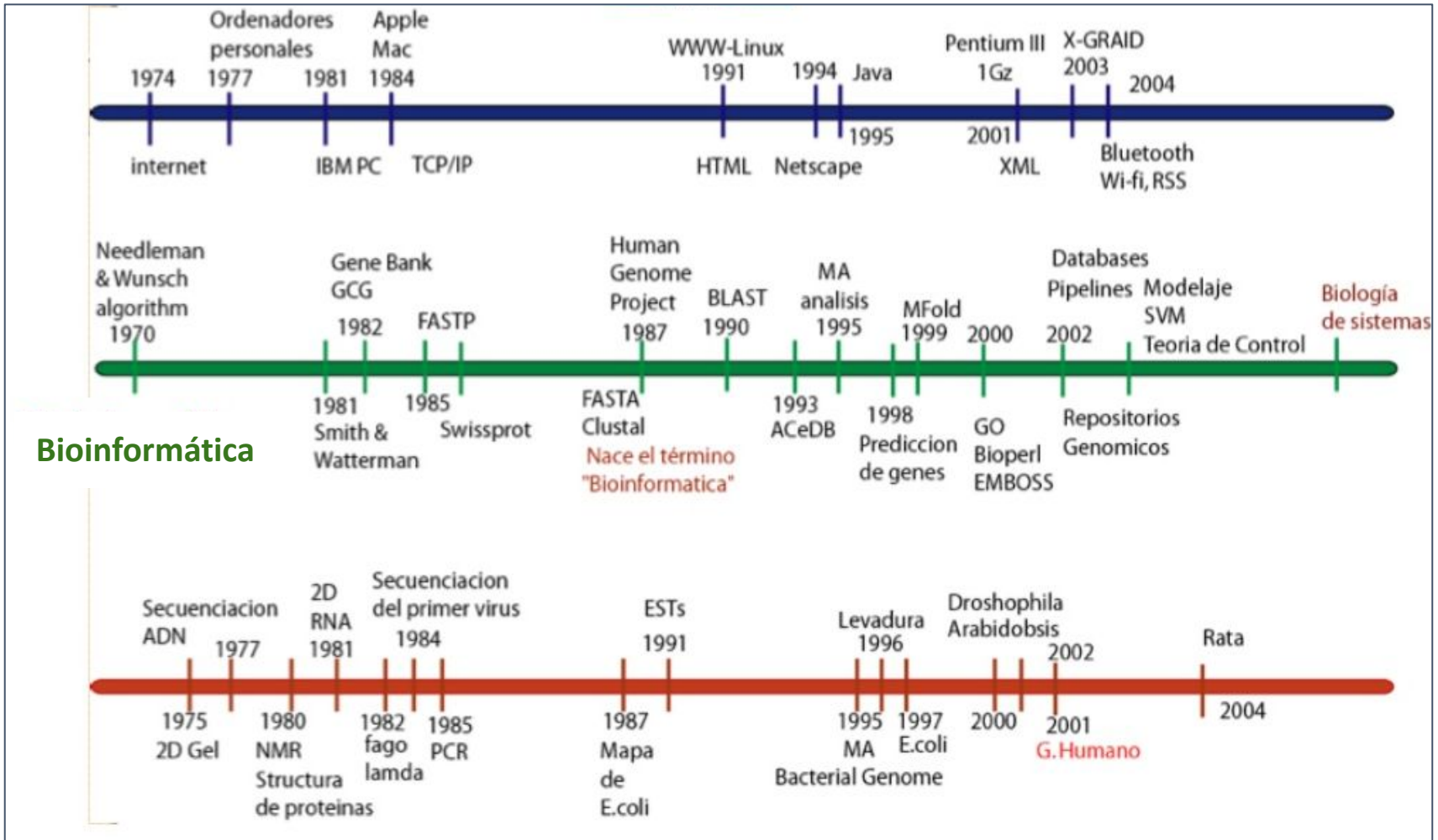


Jafari M, et al. PLoS One. 2017 Dec 21;12(12):e0189922

# Desenvolvimento da Bioinformática



# Desenvolvimento da Bioinformática



# Wellcome Sanger Institute Snapshots of our history 1992-2020

Founded to sequence the human genome, the Sanger Institute today is a biomedical research centre recognised globally for undertaking large-scale genome science that forms the foundations of knowledge in biology and medicine.



Fred Sanger, Wellcome Genome Campus

## 1993 The Sanger Centre opens

The genome sequencing facility is named after Fred Sanger, the double Nobel Laureate who devised the method for DNA sequencing used in the Human Genome Project. A modest individual, Sanger reluctantly gives permission for the use of his name. Having accepted, however, he cautions... "it better be good".

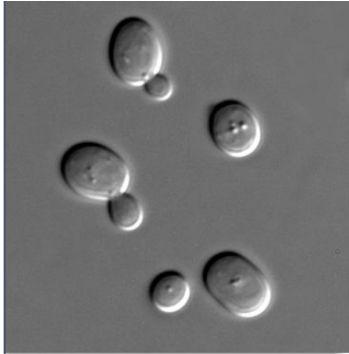


Image Credit: Masur. Licence: Public domain via Wikimedia



## 1996 The yeast genome sequence

The Sanger Centre makes the single largest contribution to sequencing the first complete genome from a eukaryotic organism (an organism whose cells have nuclei) (12,000,000 bases). *Saccharomyces cerevisiae* is a yeast used for millennia in winemaking, brewing and baking. It is also an experimental model organism widely employed to study the functions and interactions of genes. Over half of the 6,000 genes revealed by the sequence are previously unknown.



Image Credit: CDC / Alissa Eckert and James Archer



## 1998 The tuberculosis bacterium genome sequence

*Mycobacterium tuberculosis* causes tuberculosis in hundreds of millions of people worldwide, most in low- and middle-income countries. It is one of the first bacteria sequenced and the first of many pathogens sequenced at the Sanger Centre. The sequence (4,000,000 bases) contains 4,000 genes and offers new strategies for diagnosis, prevention and treatment of tuberculosis.

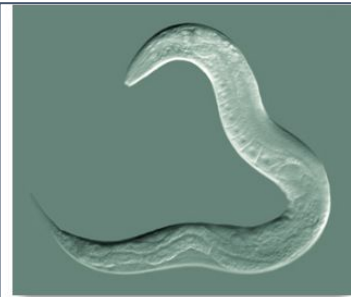


Image Credit: Bob Goldstein. Licence: CC BY-SA 3.0  
Caenorhabditis elegans

## 1998 The first animal genome sequence

*Caenorhabditis elegans* (100,000,000 bases) is a nematode worm widely used as an experimental model organism. John Sulston, the founding Director of the Sanger Centre, won the Nobel Prize for describing its developmental lineage, and the sequencing of "the worm" was fundamental in developing the tools to sequence the human genome. The sequence has enabled *C.elegans* to provide profound insights into development, neurobiology and ageing.







View of data on human chromosome 13 in Ensembl

## 2000 The Ensembl genome browser

Built as a collaboration between the Sanger Centre and EMBL's European Bioinformatics Institute (EMBL-EBI), Ensembl contains genome databases for vertebrate species. Used extensively by the global research community, it enables access to genome sequences and important additional information on genome structure and function. It continues to expand and provide services today.



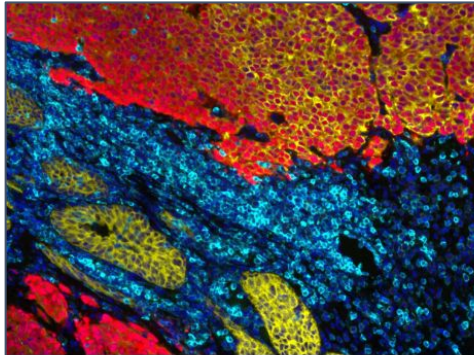
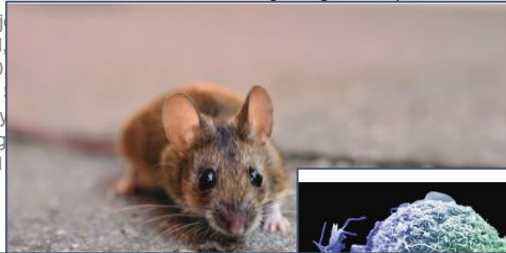


Image Credit: Mateus Crespo, The Institute of Cancer Research. Licence: CC BY 4.0

## 2000 The Cancer Genome project

The project aims to identify mutated genes. By 2020, through the Cancer Genome Atlas project, the world's systematic cancer genome project will be completed.



## 2002 The mouse genome sequence

The Sanger Institute is a leader in the International Mouse Genome Consortium.

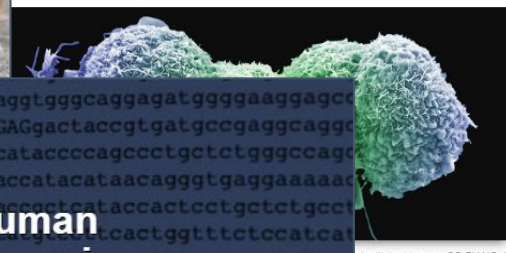


Image Credit: Sanger Institute. Licence: CC-BY-NC-4.0

## 2009 The first complete cancer genomes are sequenced

Comprehensive genome sequences of a malignant melanoma and a lung cancer by Sanger Institute researchers reveal all the mutations in cancer genomes for the first time. Analysis of the lung cancer genome suggests a typical smoker acquires one mutation in a lung cell for every 15 cigarettes smoked.



Image credit: Russ London at English Wikipedia. Licence: CC BY-SA 3.0

A printout of the human genome presented as a series of books at the Wellcome Collection, London. The 3.4 billion units of DNA code are written into more than a hundred volumes, each a thousand pages long.

## 2004 The finished human genome sequence is published

Analysis of the finished human genome is published

and culminated in the identification of new genes.

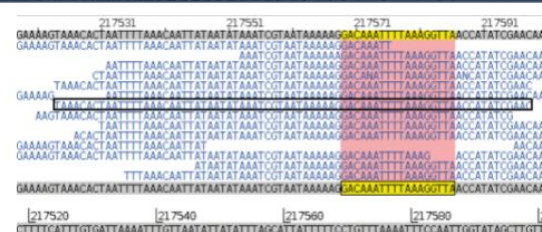
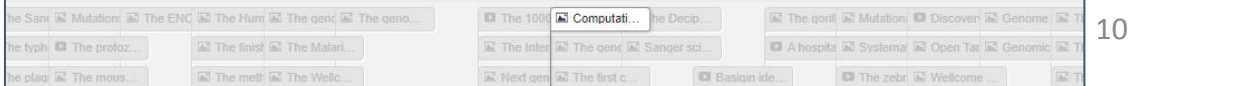


Image Credit: Adam Reid, Wellcome Sanger Institute

Reads aligned to a reference genome sequence using BWA

## 2009 Computational tools for analysing human genome sequences

Sanger researchers develop the Burrows-Wheeler Alignment (BWA) tool. The freely available software aligns short reads of DNA sequence produced by next-generation technologies against a reference sequence, such as the human genome, more accurately and faster than existing tools. BWA remains in use today and by 2020 has been cited over 17,000 times.

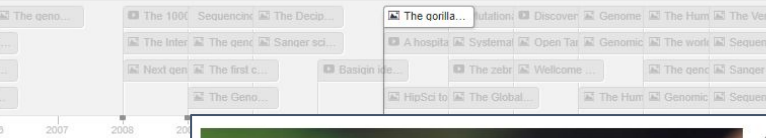




Kamilah the Gorilla  
Image Credit: San Diego Zoo

## 2012 The gorilla genome sequence

The [DNA sequence](#) (3,000,000,000 bases) of female gorilla, named Kamilah, provides insight into the evolution of the great apes, which include *sapiens*. It sheds new light on human language development, neurodegenerative illness, and information relevant to survival of mountain gorilla species.



## 2013 The zebrafish genome sequence

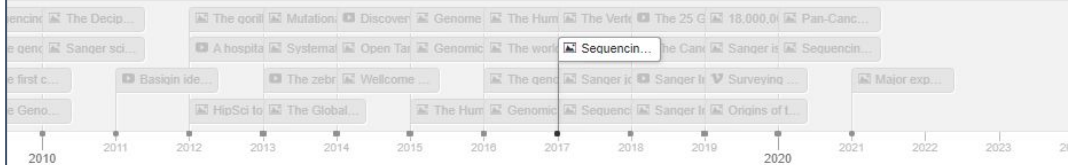
The zebrafish is an important experimental model organism used to understand embryonic development in health and disease. Sanger's [zebrafish genome sequence](#) (1,600,000,000 bases) reveals that humans share about 70 per cent of their genome with the zebrafish. Zebrafish becomes a model of choice in screening and testing new treatments for a range of diseases.



Image Credit: Pixabay

## 2017 Sequencing mosquito genomes reveals spread of insecticide resistance across Africa

Anopheles mosquitoes transmit the malaria parasites that cause hundreds of thousands of deaths every year. In the [largest study of Anopheles mosquito genomes to date](#), Sanger researchers with African collaborators discover that mosquitoes are extremely genetically diverse, providing clues as to why they quickly become insecticide-resistant.





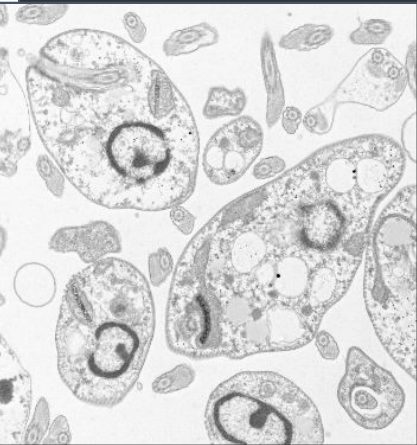
**2002**  
**The protozoan causing the deadliest form of malaria is sequenced**

*Plasmodium falciparum* each year, predominant. The genome sequence improves understanding severe disease, enables treatments, and become spread of resistance to a



**2005**  
**The genome sequence of the protozoan causing sleeping sickness**

*Trypanosoma brucei*, transmitted by tsetse flies, causes sleeping sickness, a common and deadly disease in sub-Saharan Africa. The genome



**2006**  
**The genome sequence of the protozoan causing leishmaniasis**

*Leishmania* species cause a spectrum of disfiguring and sometimes fatal diseases. Two million people in 100 countries are affected each year. Scientists at the Sanger Institute sequence 24 of the 36 chromosomes in the genome (32,800 genes) search



**2005**  
**The genome sequence of the protozoan causing amoebic dysentery**

*Entamoeba histolytica* is an intestinal parasite which causes the bowel disease amoebiasis, a significant cause of illness and death in low- and middle-income tropical countries. The genome sequence (23,700,000 bases) provided by Sanger researchers provides new information of this major

**2009**  
**The genome sequence of the worm causing schistosomiasis**

Sanger Institute researchers publish the complete sequence of *Schistosoma mansoni* (363,000,000 bases), a parasitic worm transmitted in drinking water and responsible for schistosomiasis which affects 200 million people worldwide. The genome sequence provides a foundation for understanding the parasite's complex biology enabling further research into new treatments.



Image Credit: Bruce Wetzel and Harry Schaefer, National Cancer Institute

Schistosome parasite





Image Credit: Dan Ross / Wellcome Sanger Institute

Coronavirus samples in storage at the Wellcome Sanger Institute

## 2020 Sequencing of Covid-19 genomes to monitor the pandemic at a national scale

Partners in the [COVID-19 Genomics UK consortium \(COG-UK\)](#), the Sanger Institute generates genomes at scale for national surveillance of the UK pandemic. The data enables tracking virus spread in real time, detection of new hypertransmissible strains, monitoring vaccine effectiveness and supports public health authorities in tackling local outbreaks.



Imagem: Getty Images

Em 48 horas, brasileiros  
sequenciam genoma  
do coronavírus

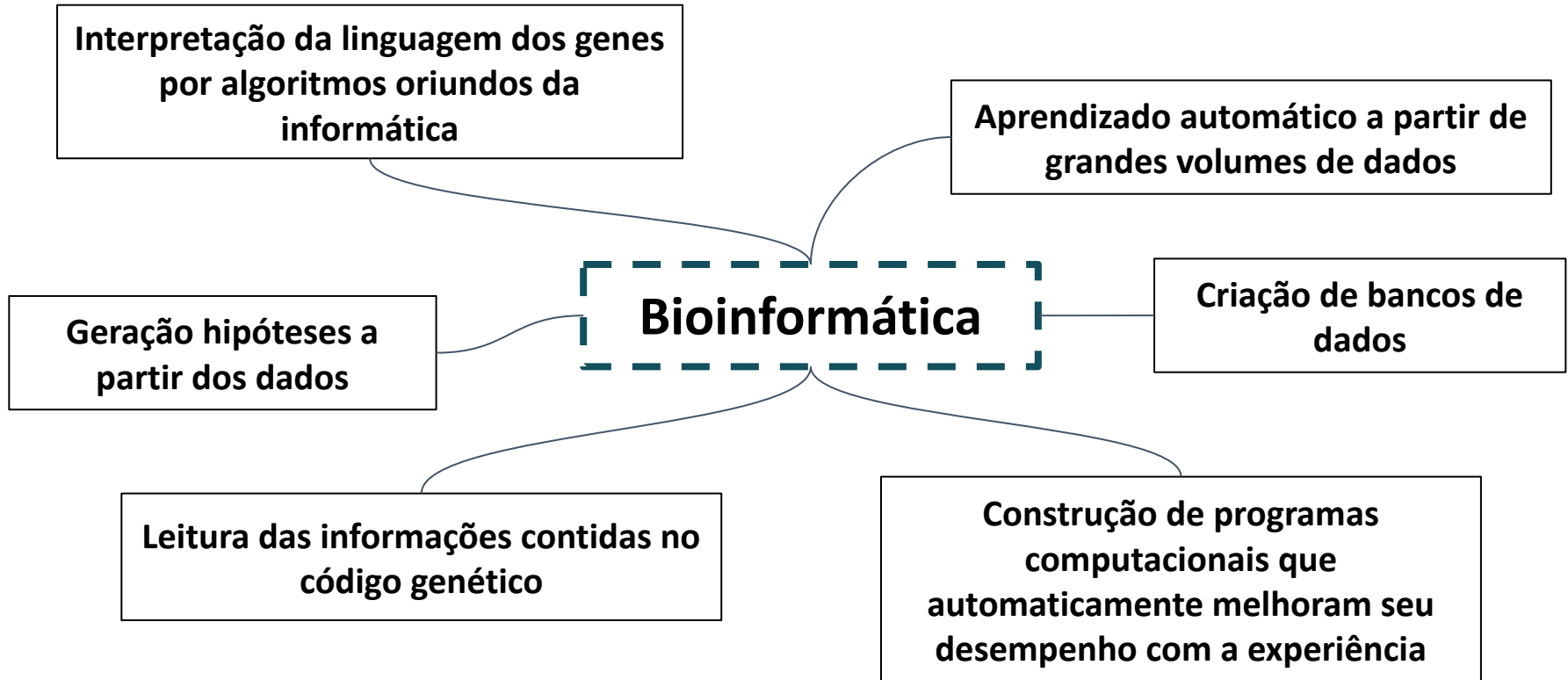
ESTADÃO conteúdo

Giovana Girardi

29/02/2020 11h20



Em só 48 horas desde a confirmação do primeiro caso no Brasil de  
pelo novo coronavírus, pesquisadores brasileiros conseguiram sequenciar  
genoma do vírus que chegou ao País. O trabalho foi conduzido por  
cientistas do Instituto Adolfo Lutz, do Instituto de Medicina Tropical da



- ❖ Ciência e tecnologia de aprendizagem, gestão e processamento de informação biológica - utilização de métodos computacionais, matemáticos e estatísticos para analisar dados biológicos, bioquímicos e biofísicos
- ❖ Biologia molecular, computação, estatística, matemática, engenharia

## Bases de dados primárias

### Nucleotídeos

GenBank (NCBI)  
ENA (European Nucleotide Archive-EMBL-EBI)  
DDBJ (DNA Bank of Japan)  
NDB (Nucleic acid database)

### Proteínas

SWISS-PROT  
UniProtKB  
PDB (Protein Data Bank)

## Bases de dados secundárias

### Proteínas

ProSite  
Pfam

### Otras

Gene-Ontology  
Refseq

## Bases de dados especializadas

ZFIN (Zebra fish)  
Flybase (*D. melanogaster*)  
TAIR (*Arabidopsis*)  
ENSEML (Homem, camundongo e outros)  
KEGG (Metabolic pathways)  
OMIM (Human genetic disorders)  
SGD (*Saccharomyces*)

# Entendendo as ferramentas!!!

## BLAST

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

### Basic Local Alignment Search Tool

**BLAST** finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)


**NEWS**

BLAST+ 2.10.1 is released – Fix for TBLASTN Multi-Threading issue.

This version supports pulling databases from our FTP site as well from [cloud providers](#) or our [BLAST+Docker solution](#).

Thu, 18 June 2020 12:00:00 EST [More BLAST news...](#)

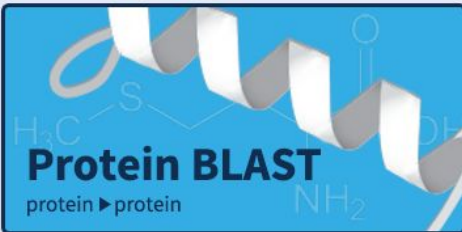
### Web BLAST



**Nucleotide BLAST**  
nucleotide ▶ nucleotide

**blastx**  
translated nucleotide ▶ protein

**tblastn**  
protein ▶ translated nucleotide

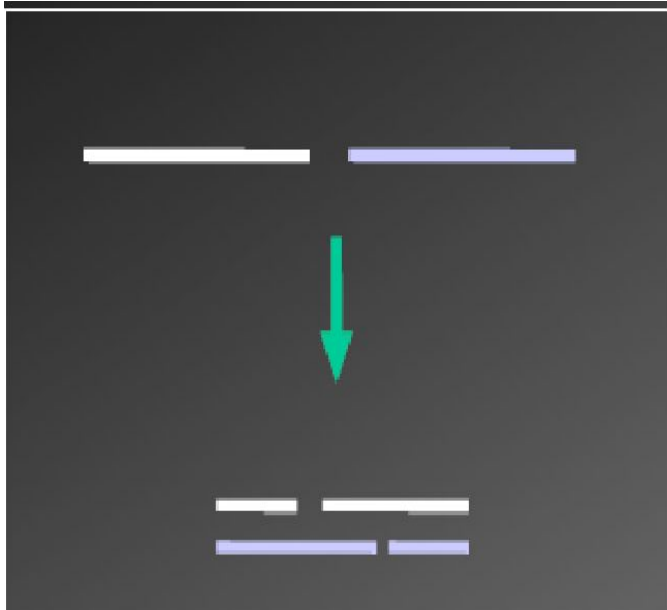


**Protein BLAST**  
protein ▶ protein

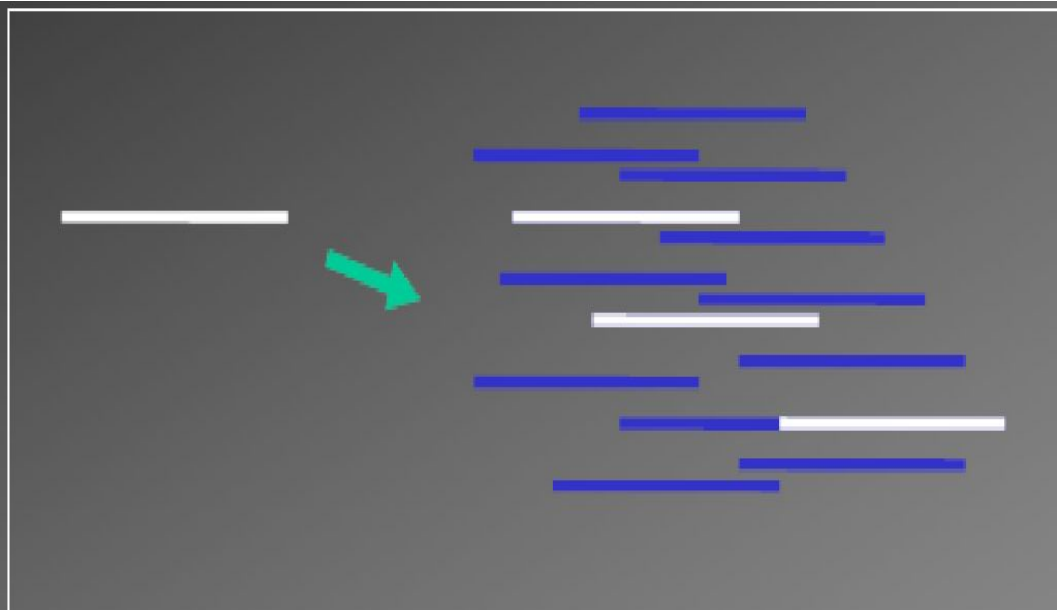
### BLAST Genomes

[Human](#)   [Mouse](#)   [Rat](#)   [Microbes](#)





**Alinhamento de seqüências**



**Busca de sítios homólogos**

- ❖ Cadeias curtas de alta similaridade
- ❖ Grandes bancos de dados
- ❖ E-value  $< 1 e^{-05}$  → muito confiável

Homólogos → origem comum em organismos diferentes. Não indica a mesma função

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download ▾

Manage Columns ▾

Show 100 ▾ 

select all 100 sequences selected

[GenPept](#) [Graphics](#)

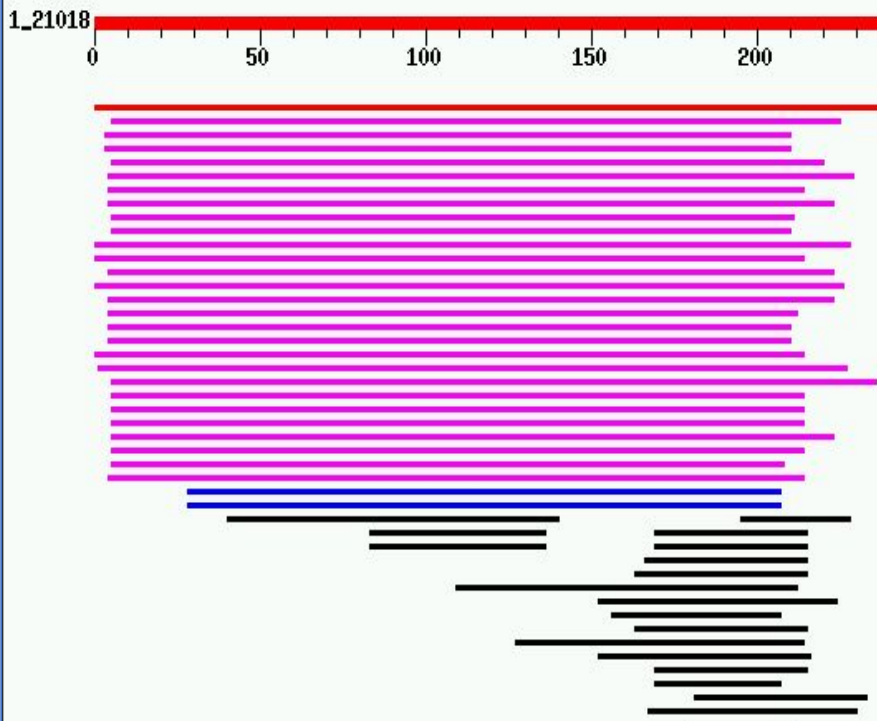
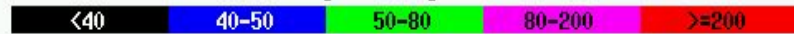
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase [Mus musculus]</a>	148	397	19%	8e-33	100.00%	<a href="#">NP_034531.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase [Mus caroli]</a>	147	396	19%	2e-32	98.78%	<a href="#">XP_021029505.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase [Mastomys coucha]</a>	143	389	19%	3e-31	96.34%	<a href="#">XP_031204984.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase [Arvicanthus niloticus]</a>	143	387	19%	3e-31	96.34%	<a href="#">XP_034343177.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase [Meriones unguiculatus]</a>	143	380	19%	4e-31	95.12%	<a href="#">XP_021511140.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase [Grammomys surdaster]</a>	142	382	19%	5e-31	96.34%	<a href="#">XP_028638877.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase [Rattus norvegicus]</a>	142	387	19%	9e-31	95.12%	<a href="#">NP_058855.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase [Cricetulus griseus]</a>	141	380	19%	1e-30	95.12%	<a href="#">XP_003501465.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: histidine ammonia-lyase [Peromyscus maniculatus bairdii]</a>	140	379	19%	3e-30	93.90%	<a href="#">XP_006988520.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase [Mus pahari]</a>	140	387	19%	3e-30	95.12%	<a href="#">XP_021060879.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase [Nannospalax galili]</a>	140	370	19%	4e-30	91.46%	<a href="#">XP_008830287.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase isoform X2 [Leptonychotes weddellii]</a>	132	354	19%	4e-30	86.59%	<a href="#">XP_030893358.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase [Mesocricetus auratus]</a>	139	378	19%	6e-30	92.68%	<a href="#">XP_005068059.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase isoform X1 [Leptonychotes weddellii]</a>	132	353	19%	1e-29	86.59%	<a href="#">XP_030893357.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase [Microtus ochrogaster]</a>	138	378	19%	1e-29	92.68%	<a href="#">XP_005358216.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase isoform X2 [Pan troglodytes]</a>	137	365	19%	1e-29	91.46%	<a href="#">XP_024203778.1</a>
<input checked="" type="checkbox"/>	<a href="#">histidine ammonia-lyase isoform X2 [Nomascus leucogenys]</a>	137	362	19%	2e-29	91.46%	<a href="#">XP_024203778.1</a>

 Feed

### Distribution of 50 Blast Hits on the Query Sequence

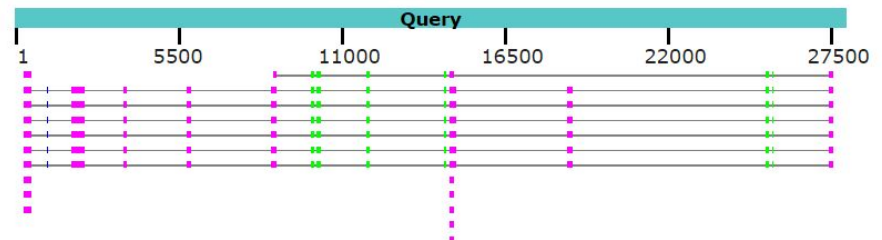
Mouse-over to show define and scores. Click to show alignments

#### Color Key for Alignment Scores



Alignment Scores ■ < 40 ■ 40 - 50 ■ 50 - 80 ■ 80 - 200 ■ >= 200

### Distribution of the top 200 Blast Hits on 100 subject sequences



[Download](#) [GenPept](#) [Graphics](#) Sort by: E value

#### histidine ammonia-lyase-like [Leptonychotes weddellii]

Sequence ID: [XP\\_030895515.1](#) Length: 384 Number of Matches: 9

Range 1: 167 to 239 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Pre

	Score	Expect	Method	Identities	Positives	Gaps	Frame
	126 bits(317)	3e-33	Compositional matrix adjust.	63/73(86%)	65/73(89%)	5/73(6%)	+2
Query	14567		ALDYLAIGVHELAAISERRIERLCNPSELPAFLVAEAGGLNSGFMIHCTAAALGKDT-				14743
Sbjct	167		ALDYLAIGVHELAAISERRIERLCNPSELPAFLVAEAGGLNSGFMIHCTAAALGK+T				226
Query	14744		----PLVRTQACP	14770			
			P RTQACP				
Sbjct	227		QEYCPFPRTQACP	239			

# Clustal omega

<https://www.ebi.ac.uk/Tools/msa/clustalo/>

Input form | Web services | Help & Documentation | Bioinformatics Tools FAQ | Feedback | Share

## Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

**Important note:** This tool can align up to 4000 sequences or a maximum file size of 4 MB.

**STEP 1 - Enter your input sequences**

Enter or paste a set of

PROTEIN

sequences in any supported format:

Or, upload a file: [Seleccionar archivo](#) Ningún archivo seleccionado [Use a example sequence](#) | [Clear sequence](#) | [See more example inputs](#)

**STEP 2 - Set your parameters**

OUTPUT FORMAT

ClustalW with character counts

The default settings will fulfill the needs of most users.

[More options...](#) (Click here, if you want to view or change the default settings.)

**STEP 3 - Submit your job**

Be notified by email (Tick this box if you want to be notified by email when the results are available)

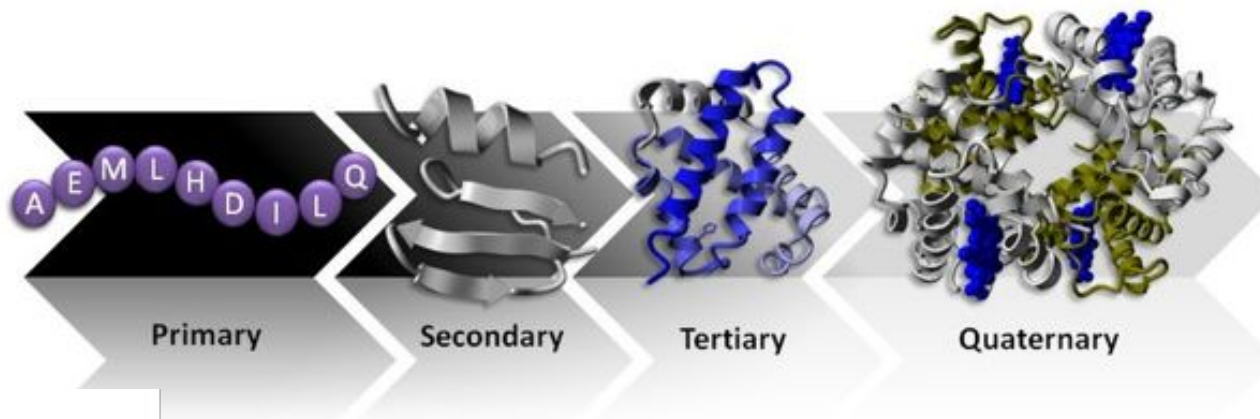
Submit

G.gallus	MPRYTVHVRGEHLAVPCPHGTNTVGWLGKEAVRRYMKNKPDNGGFTSVEEVKFFVRRCKG	60
M.musculus	MPRYTVHVRGEHLAVPCQDGKLTVGWLGREAVRRYMKNKPDNGGFTSVDEVQFLVHRCCKG	60
T.cruzi	-----	0
T.grayi	-----	0
G.gallus	LGLLDLDDTVEDALEDNFVVEVIEGDIMSPDFIPSQPEGVHLYSKYREPEQYISLDGNS	120
M.musculus	LGLLDNEDELEVALEDNFVVEVIEGDVMSPDFIPSQPEGVFLYSKYREPEKYIALDGDS	120
T.cruzi	-----MRVILDGCS	9
T.grayi	-----MKVTLDGCS	9
	: *** *	
G.gallus	LTTQDLVNLGKGL-YKIKLTPAEAKVKQSREVIERIVKEQTWVYGITTFGFKFARTVIP	179
M.musculus	LSTEDLVNLGKGR-YKIKLTSIAEKKVQQSREVIDSIIKERTWVYGITTFGFKFARTVIP	179
T.cruzi	LTPDVLALGYEKGATIEISDEAVARITAARAVIDKIVNDRQTVYGINTEGFKFESTIIP	69
T.grayi	LTPDALYALGHEKDISTILLAEAAIRRRINAGRAVIDKIVNERQTVYGINTEGFKFESTVIP	69
	*: : * ** . * :: * :: . * **: *::: .****.***** *:**	
G.gallus	NSKLMELQMNLVRSHSAGVGKPLTPERSRMLLALRINVLAKGYSGISLETLQQVIEAFNA	239
M.musculus	ANKLQELQVNLVRSHSSGVGKPLSPERCRLMLLALRINVLAKGYSGISLETLKQVIEAFNA	239
T.cruzi	PHQLEELQLNLRSHSACVGEPLTPERARMMLALRVNVLCKGHSGIRLETVQKVVKAFNA	129
T.grayi	PNQLVDLQLNLRSHSACVGEPLKPKQRARMMMLALRINILCKGHSGIRLETVEKVVKAFNA	129
	:* :*:*:*:****: **:**:*.**:*:*:*:*:*:*:*:*:** ***:::: :****	
G.gallus	SCLPYIPEKGTVGASGDLAPLSHLALGLTGEGKMWSPK-SGNADAKYVLEAHGLKPITLK	298
M.musculus	SCLSYVPEKGTVGASGDLAPLSHLALGLIGEGKMWSPK-SGNADAKYVLEAHGLKPIVLK	298
T.cruzi	GVPYIPEQGTVGASGDLGPLSHLALGMLGEGLLATLNKKFRDAGSVLRELGVPEITLA	189
T.grayi	GVPPIPEQGTVGASGDLGPLSHLALGMLGEGMLATLNKPVFRDASTVRELGVPEITLA	189
	. : :*:*:*****.*****: ** : : : : ** ** . *:***.*	
G.gallus	PKEGLALINGTQMITSLGCEAVERASAIARQADIVAALTLEVLKGTTKAFDTDIHAVRPH	358
M.musculus	PKEGLALINGTQMITSLGCEALERASAIARQADIVAALTLEVLKGTTKAFDTDIHAVRPH	358
T.cruzi	AKEGLALINGTQFISALGAEAVVRARKIARLADVALAMSHEALRATNSTLNPDIHVRPH	249
T.grayi	AKEGLALINGTQFMSALGAEAVVRARRIARLADVLSMSHEALQSTVSSLNPDIHVRPH	249
	*****:***:**: ** ** ** **: : : *.*:.* :::: ** ****	
G.gallus	PGQAFVADPDRCHLRSDVNDGFTLEGHREGRVDRANTVDRGDRMLGAVDRITLAVDRM	419



<https://www.rcsb.org>

1971 Criação do PDB.  
Em 1974 tinha 12 estruturas  
24/06/21 → 179210



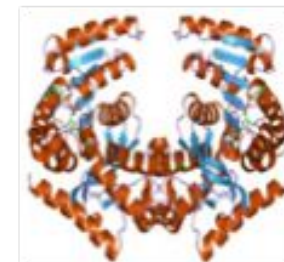
myoglobin



hemoglobin



carboxypeptidase A



lactate dehydrogenase

# Uniprot

<https://www.uniprot.org>

# Pfam

<https://pfam.xfam.org>

Source	Domain	Start	End
colled_coll		1/3	42/62
Pfam	Acyl-CoA_dh_H	46	160
Pfam	Acyl-CoA_dh_H	104	261
Pfam	Acyl-CoA_dh_1	273	421

# Modelagem de proteínas por homologia

The screenshot shows the SWISS-MODEL web interface. At the top left, there is a logo for BIOZENTRUM SIB, University of Basel, The Center for Molecular Life Sciences. The main header is "SWISS-MODEL". To the right of the header are navigation links: "Modelling", "Repository", "Tools", "Documentation", "Log in", and "Create".

The main content area is titled "Start a New Modelling Project". It contains a form for entering a target sequence. The form has a text input field with the placeholder text "Paste your target sequence(s) or UniProtKB AC here". Below the input field are two buttons: a green button labeled "+ Upload Target Sequence File..." and a grey button labeled "Validate".

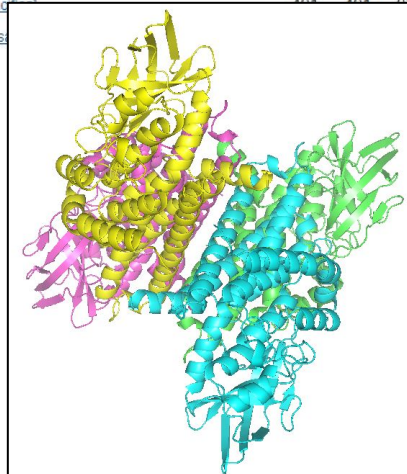
Below the input field, there are three more input fields: "Project Title:" with the value "Untitled Project", "Email:" with the value "Optional", and a "Search For Templates" button. To the right of the form is a "Supported Inputs" section with a dropdown menu showing "Sequence(s)", "Target-Template Alignment", "User Template", and "DeepView Project".

At the bottom of the form, there are two large blue buttons: "Search For Templates" and "Build Model".

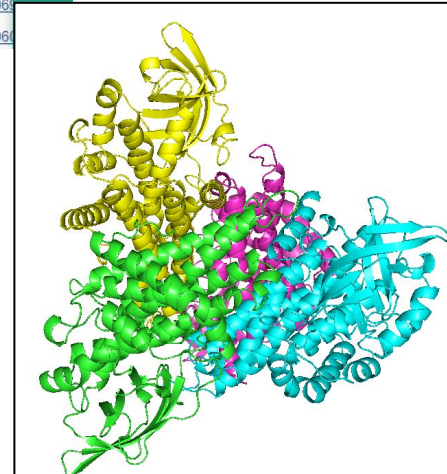
<https://swissmodel.expasy.org>



	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> isovaleryl-coA dehydrogenase [Trypanosoma rangeli SC58]	707	707	98%	0.0	83%	<a href="#">ESL07937.1</a>
<input type="checkbox"/> isovaleryl-coA dehydrogenase [Trypanosoma theileri]	654	654	92%	0.0	81%	<a href="#">ORC91678.1</a>
<input type="checkbox"/> isovaleryl-coA dehydrogenase_putative [Trypanosoma brucei gambiense DAL972]	627	627	94%	0.0	76%	<a href="#">XP_011779312.1</a>
<input type="checkbox"/> isovaleryl-coA dehydrogenase_putative [Trypanosoma brucei brucei TREU927]	626	626	94%	0.0	76%	<a href="#">XP_828243.1</a>
<input type="checkbox"/> unnamed protein product [Trypanosoma congolense IL3000]	595	595	98%	0.0	72%	<a href="#">CCD16413.1</a>
<input type="checkbox"/> putative isovaleryl-coA dehydrogenase [Leishmania infantum JPCM5]	580	580	99%	0.0	71%	<a href="#">XP_001466486.1</a>
<input type="checkbox"/> putative isovaleryl-coA dehydrogenase [Leishmania major strain Friedlin]	579	579	95%	0.0	71%	<a href="#">XP_003721899.1</a>
<input type="checkbox"/> putative mitochondrial isovaleryl-coA dehydrogenase [Leptomonas pyrrhocoris]	579	579	96%	0.0	71%	<a href="#">XP_015659347.1</a>
<input type="checkbox"/> putative isovaleryl-coA dehydrogenase [Leishmania mexicana MHOM/GT/2001/U1103]	577	577	98%	0.0	71%	<a href="#">XP_003876667.1</a>
<input type="checkbox"/> putative isovaleryl-coA dehydrogenase [Leishmania braziliensis MHOM/BR/75/M2904]	575	575	99%	0.0	70%	<a href="#">XP_001565860.1</a>
<input type="checkbox"/> isovaleryl-coA dehydrogenase_putative [Leishmania panamensis]	575	575	95%	0.0	71%	<a href="#">XP_010700182.1</a>
<input type="checkbox"/> isovaleryl-coA dehydrogenase_putative [Leishmania guyanensis]	573	573	99%	0.0	70%	<a href="#">CCM16690.1</a>
<input type="checkbox"/> isovaleryl-coA dehydrogenase_putative [Bodo saltans]	573	573	94%	0.0	69%	<a href="#">CUE72386.1</a>
<input type="checkbox"/> putative isovaleryl-coA dehydrogenase [Leptomonas seymouri]	566	566	91%	0.0	72%	<a href="#">KPI87702.1</a>
<input type="checkbox"/> putative isovaleryl-coA dehydrogenase_fragment [Trypanosoma vivax Y486]	554	554	85%	0.0	78%	<a href="#">CCC52663.1</a>
<input type="checkbox"/> isovaleryl-CoA dehydrogenase [Sandaracinus amylolyticus]	496	496	91%	2e-172	64%	<a href="#">WP_053234518.1</a>
<input type="checkbox"/> isovaleryl-CoA dehydrogenase [Nannocystis exedens]	494	494	90%	1e-171	64%	<a href="#">WP_096327858.1</a>
<input type="checkbox"/> isovaleryl-CoA dehydrogenase [Plesiocystis pacifica]	464	464	89%	3e-170	62%	<a href="#">WP_006111111.1</a>
<input type="checkbox"/> isovaleryl-CoA dehydrogenase [Enhygromyxa sp.]	404	404	87%	2e-165	62%	<a href="#">WP_106111111.1</a>



Model 1

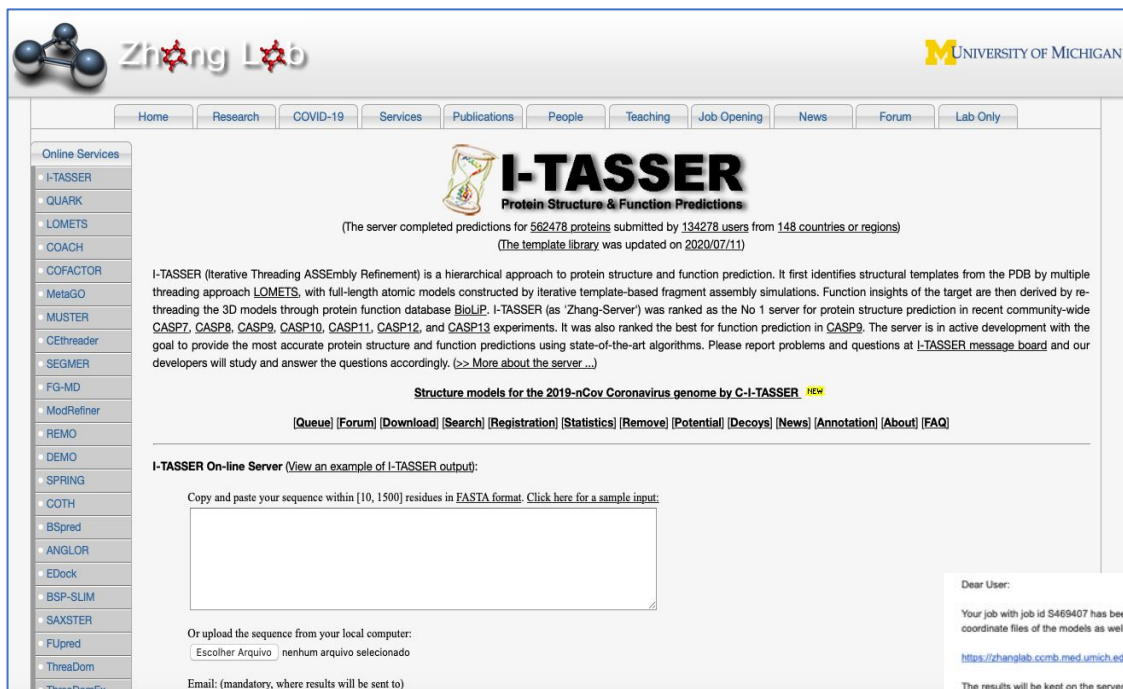


Model 2

Model 1  1jqi: Crystal Structure of Rat Short Chain Acyl-CoA Dehydrogenase Complexed With Acetoacetyl-CoA

Model 2  ivh: Structure of human isovaleryl-coa dehydrogenase at 2.6 angstroms resolution: structural basis for substrate specificity

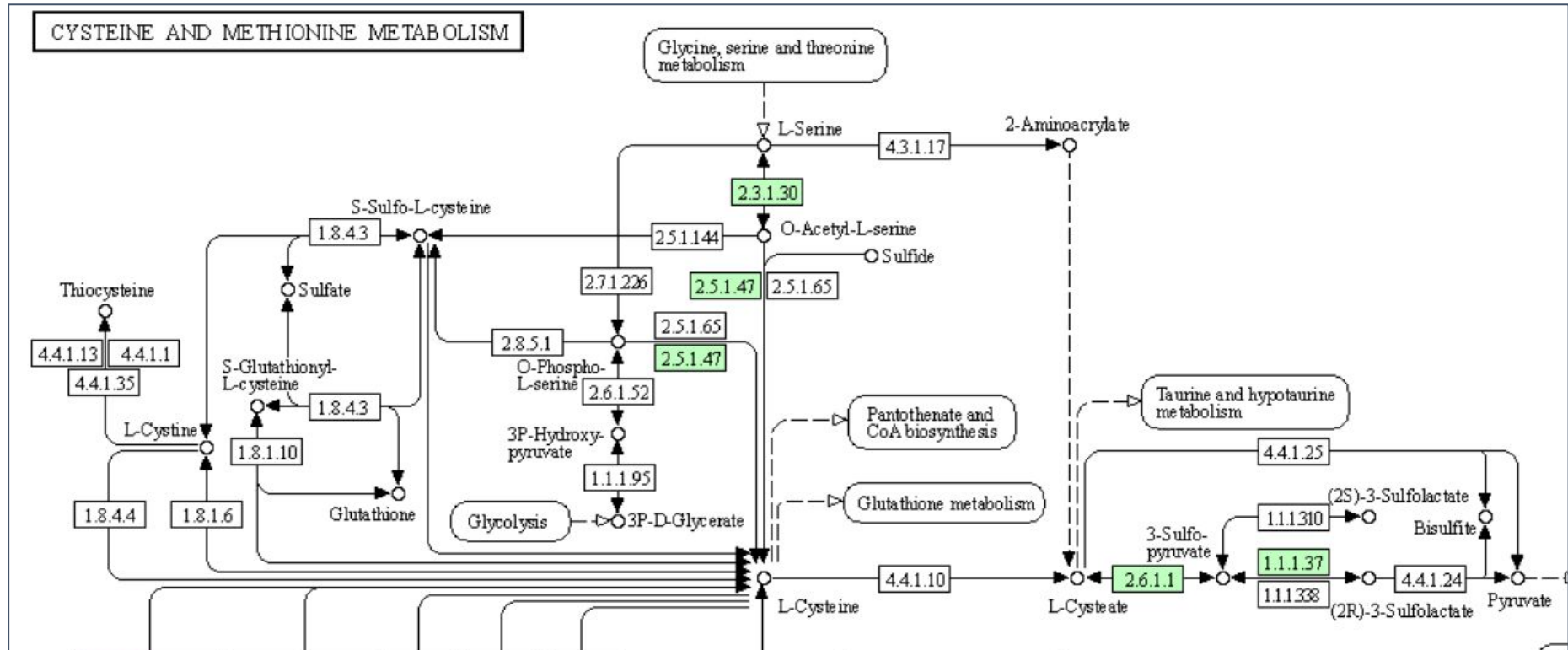
# Modelagem de proteínas por homologia



The screenshot shows the I-TASSER web interface. At the top, there are logos for 'Zhang Lab' and 'UNIVERSITY OF MICHIGAN'. Below the logos is a navigation menu with options: Home, Research, COVID-19, Services, Publications, People, Teaching, Job Opening, News, Forum, and Lab Only. On the left side, there is a vertical menu titled 'Online Services' with various options like I-TASSER, QUARK, LOMETS, COACH, COFACTOR, MetaGO, MUSTER, CEThreader, SEGMENT, FG-MD, ModRefiner, REMO, DEMO, SPRING, COTH, BSpred, ANGLOR, EDock, BSP-SLIM, SAXSTER, FUpred, ThreaDom, and ThreaDomFy. The main content area features the I-TASSER logo and the text 'Protein Structure & Function Predictions'. Below this, it states: '(The server completed predictions for 562478 proteins submitted by 134278 users from 148 countries or regions) (The template library was updated on 2020/07/11)'. A paragraph describes the I-TASSER approach: 'I-TASSER (Iterative Threading ASSEMBLY Refinement) is a hierarchical approach to protein structure and function prediction. It first identifies structural templates from the PDB by multiple threading approach LOMETS, with full-length atomic models constructed by iterative template-based fragment assembly simulations. Function insights of the target are then derived by re-threading the 3D models through protein function database BioLiP. I-TASSER (as 'Zhang-Server') was ranked as the No 1 server for protein structure prediction in recent community-wide CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, and CASP13 experiments. It was also ranked the best for function prediction in CASP9. The server is in active development with the goal to provide the most accurate protein structure and function predictions using state-of-the-art algorithms. Please report problems and questions at I-TASSER message board and our developers will study and answer the questions accordingly. (>> More about the server...)' A link is provided: 'Structure models for the 2019-nCov Coronavirus genome by C-I-TASSER NEW'. Below this are links for Queue, Forum, Download, Search, Registration, Statistics, Remove, Potential, Decoys, News, Annotation, About, and FAQ. The 'I-TASSER On-line Server' section includes a note: '(View an example of I-TASSER output):' and a text area for 'Copy and paste your sequence within [10, 1500] residues in FASTA format. Click here for a sample input:'. There is also an option to 'Or upload the sequence from your local computer:' with a file selection button and an email field for 'Email: (mandatory, where results will be sent to)'. On the right side of the screenshot, there is a 'Dear User:' message: 'Your job with job id S469407 has been completed on the I-TASSER server. The picture of the predicted models is attached with this mail. The complete results including coordinate files of the models as well as function predictions are available at: https://zhanglab.ccmb.med.umich.edu/I-TASSER/output/S469407/ The results will be kept on the server for 2 months. Thanks for using the I-TASSER server. - The I-TASSER Server Team Department of Computational Medicine and Bioinformatics University of Michigan http://zhanglab.ccmb.med.umich.edu/I-TASSER http://zhanglab.ccmb.med.umich.edu/bbs/?c=forum/2' At the bottom right of the screenshot, there are five small thumbnail images of protein structures, labeled '5 anexos'.

# Bases de dados reconstruções metabólicas

## Metabolismo da Cisteína: KEGG (*T. cruzi*)



# Bases de dados + evidencia experimental → Construção de modelos metabólicos

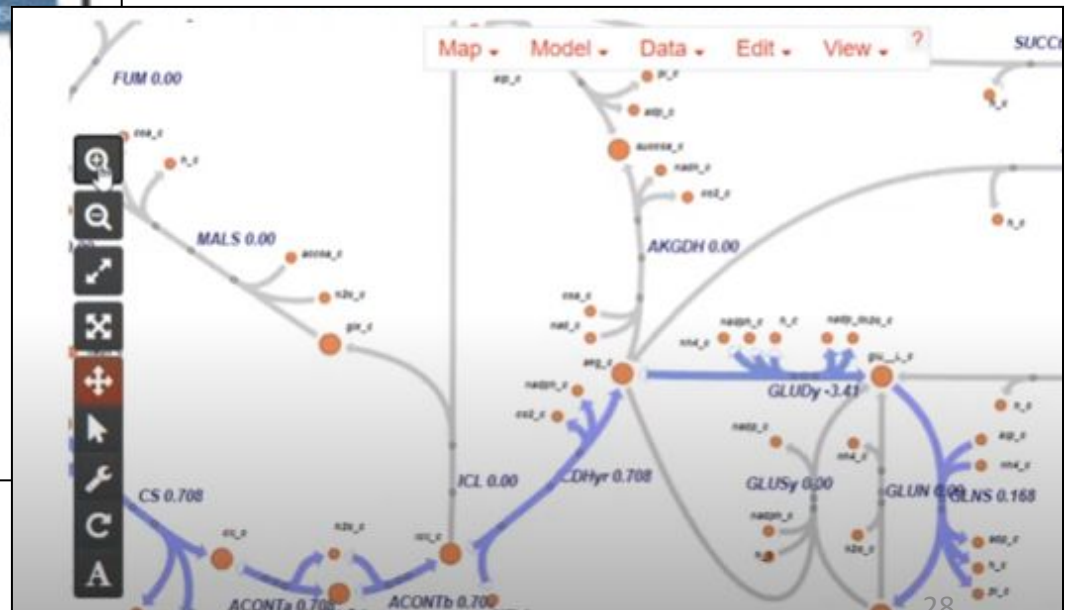
A Classic reconstruction workflow



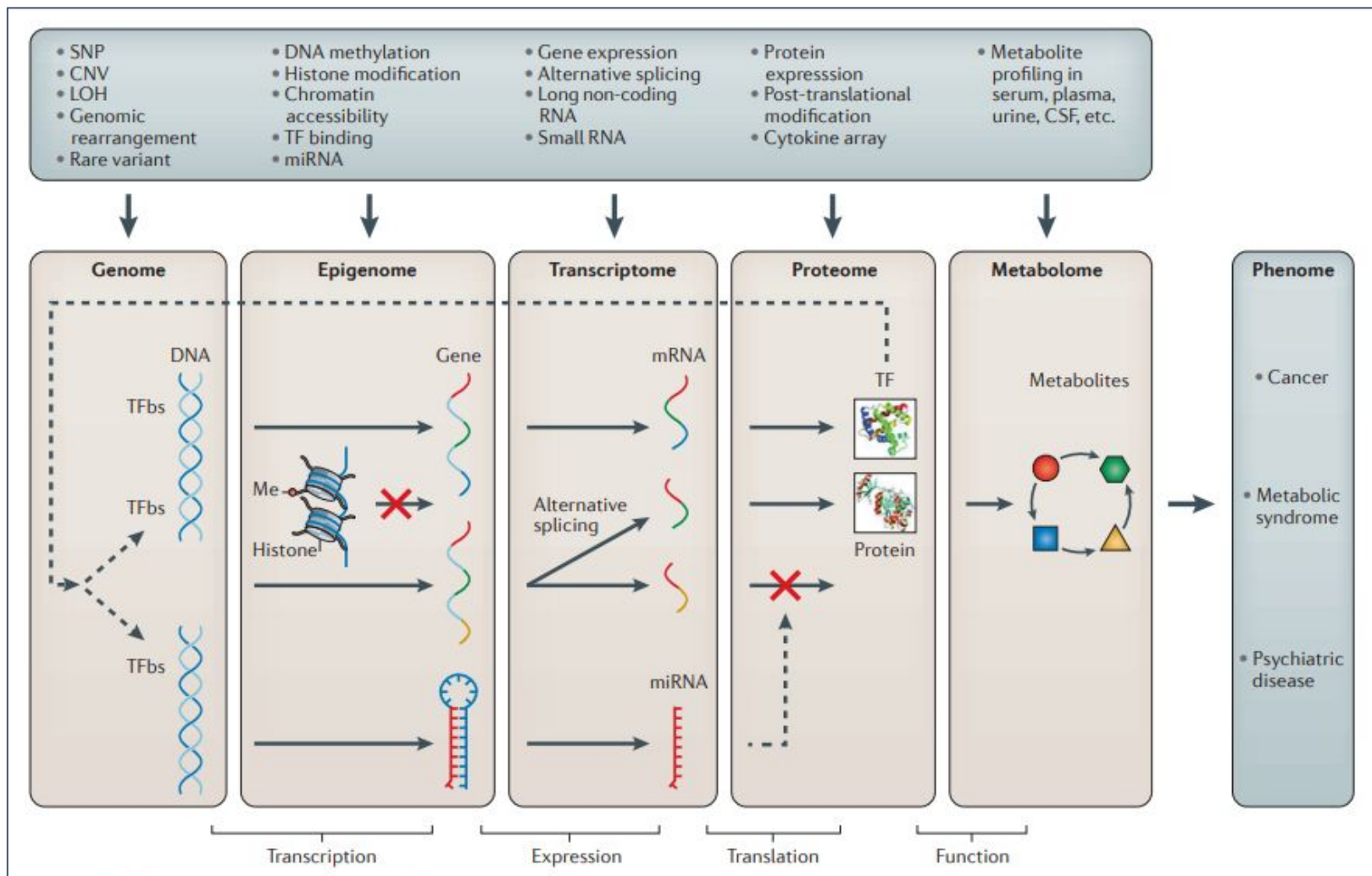
- Modelo universal:
- 2383 metabólitos (1503 únicos)
- 4383 reações, das quais:
  - 2463 enzimáticas
  - 1380 transportes



BioWinformática Tutoriais

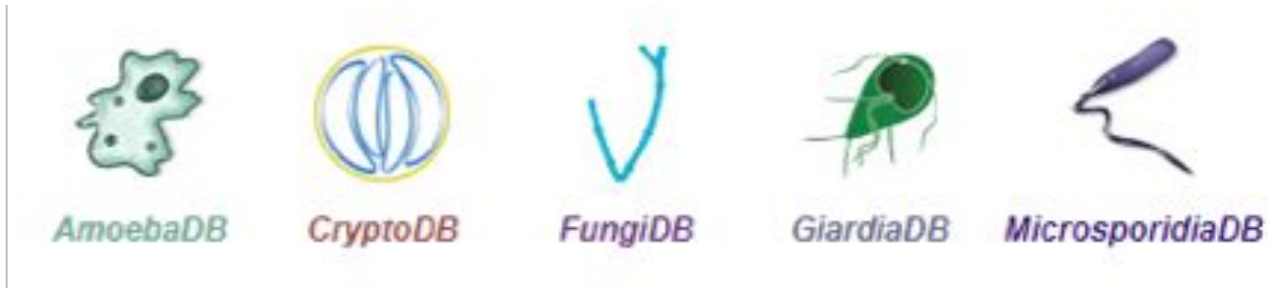


# Ômicas





TRUQUES QUE FACILITAM  
A NOSSA VIDA



## Exemplos para pesquisa em bases de dados:

Nome	Abrev.	NCBI (GenBank)	Uniprot	PDB	EC	TriTrypsDB (TcCLB)	TriTrypsDB (Dm28)
Histidine Ammonia Lyase	HAL	ESS65584	V5BD20 V5BD20_TRYCR	<u>6V6H</u>	4.3.1.3	TcCLB.506247.220	BCY84_07468
Serine AcetylTransferase	SAT	RNC54559	Q4DJX5 Q4DJX5_TRYCC	<u>4H7O*</u>	2.3.1.30	TcCLB.510879.80	BCY84_07439
Enoyl-CoA Hydratase	ECH	ESS70836	Q4E679 Q4E679_TRYCC	<u>5KJP*</u>	4.2.1.17	TcCLB.508153.130	C4B63_20g212
Cysteine Synthase	CS	XP_805193	Q4CST7 Q4CST7_TRYCC	<u>4AEC*</u>	2.5.1.47	TcCLB.507165.50	TCDM_02303
Isovaleryl-CoA dehydrogenase	IVDH	PBJ72985	Q4DQC2 Q4DQC2_TRYCC	<u>4O5M*</u>	1.3.8.4	TcCLB.506629.220	TCDM_08009

**Vamos usar/provar as ferramentas  
para responder as dúvidas!!!**