

BIOINFORMÁTICA e AULA PRÁTICA

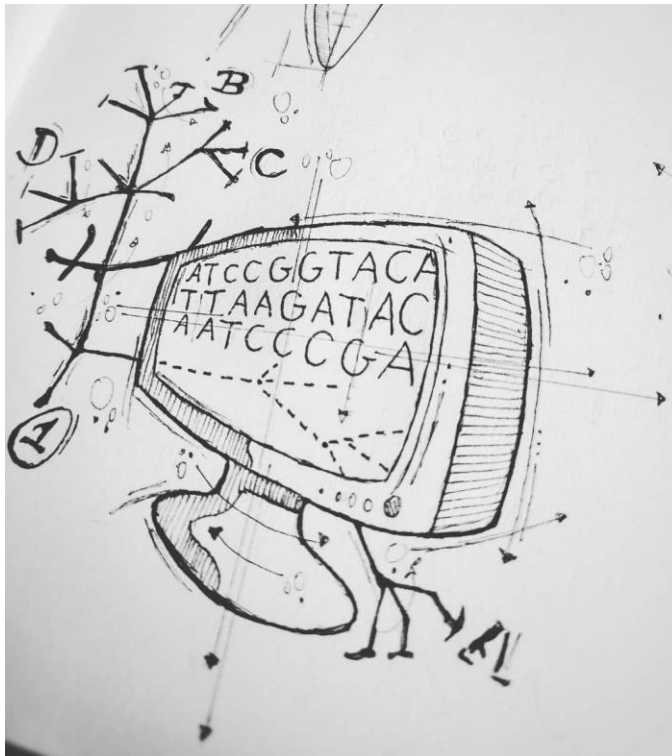
SOBRE DADOS BIOLÓGICOS:

CONHECENDO O NCBI

Aula 12

LGN232 – Genética Molecular

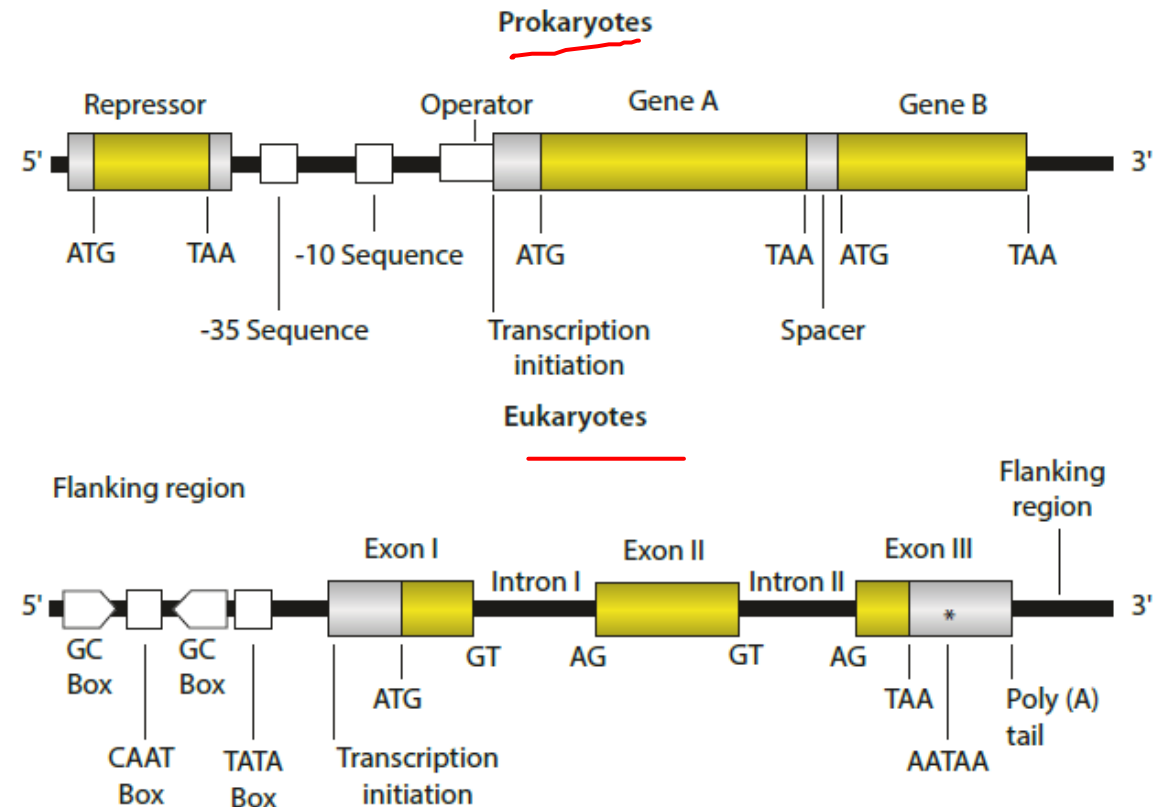
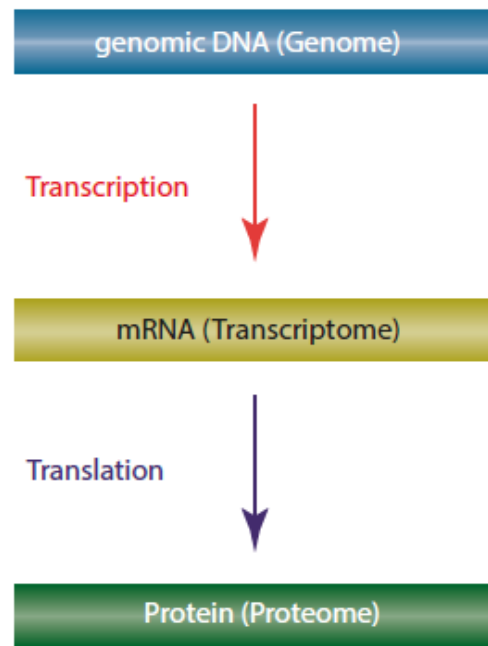
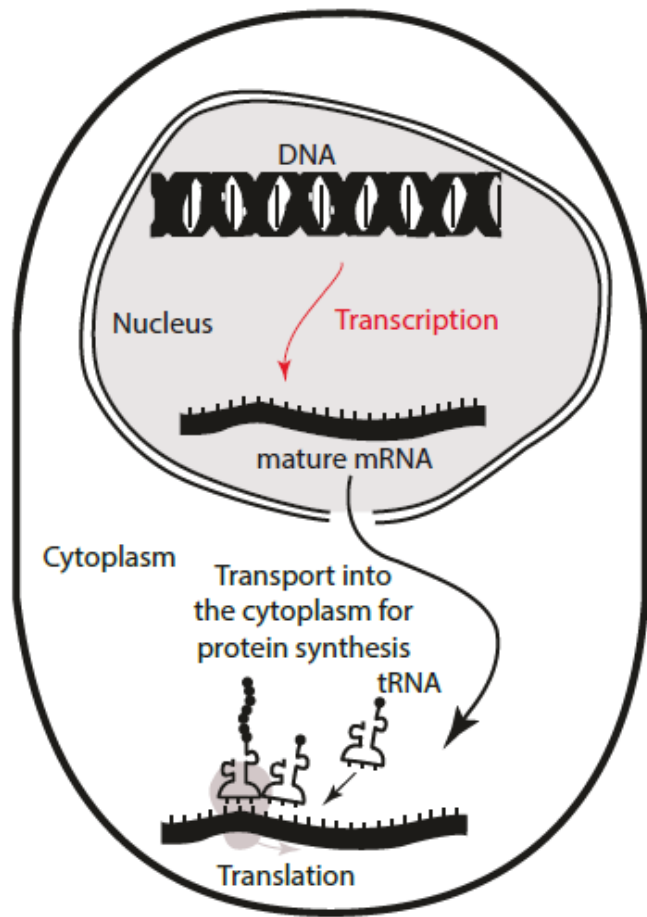
Emiliana Manesco Romagnoli
Departamento de Genética
emilianaromagnoli@usp.br



O plano para hoje >>>>>>>>>>

1. O que é Bioinformática?
2. Bancos de dados em biologia: foco em alguns bancos de dados e ferramentas do NCBI





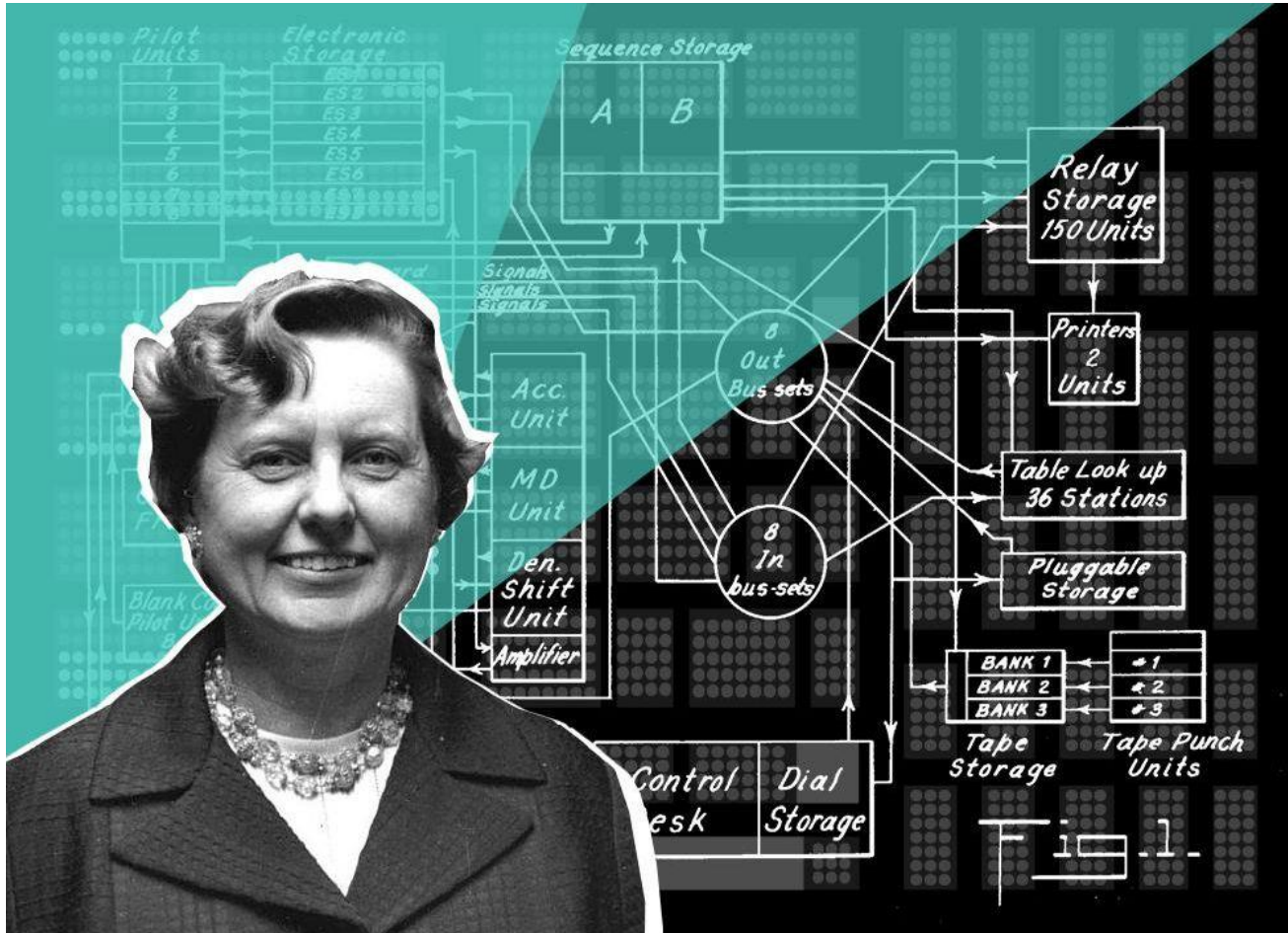
O fluxo de informações sempre procede do genoma para o proteoma, e não vice-versa. As exceções são reações que são catalisadas pela transcriptase reversa e replicase de vírus de RNA



- Momento chave para a bioinformática o início da década de 1950, quando a revista *Nature* publicou o trabalho clássico sobre a estrutura em hélice da molécula de DNA por James Watson e Francis Crick.

As bases moleculares para o entendimento estrutural da replicação e tradução do material genético foram apresentadas, permitindo-nos entender como aquela "sequência de letras" (as bases do DNA) se organizam tridimensionalmente.

Considerada a mãe da bioinformática

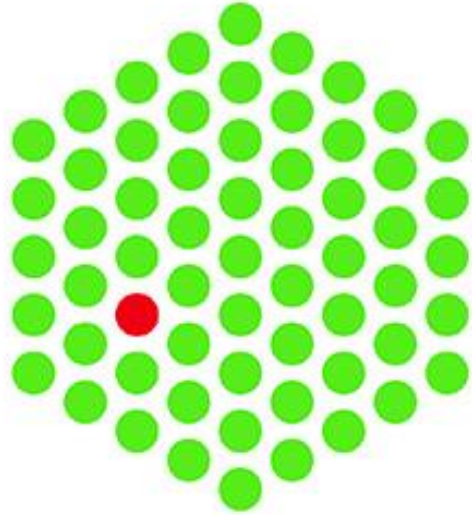


A Dr. **Margaret Belle (Oakley) Dayhoff** (11 de março de 1925 – 5 de fevereiro de 1983) foi uma físico-química americana e pioneira no campo da bioinformática. Dayhoff foi professora no Centro Médico da Universidade de Georgetown e pesquisadora em bioquímica na National Biomedical Research Foundation, onde foi pioneira na aplicação de métodos matemáticos e computacionais ao campo da bioquímica. Ela dedicou sua carreira à aplicação das tecnologias computacionais para promover avanços no campo da biologia e da medicina, em especial à criação de bases de dados de ácidos nucleicos e proteínas, além de ferramentas para utilizá-las.



IBM 7090, computador que Margaret Dayhoff utilizou no início de seus trabalhos (NASA Ames Resarch Center, 1961).

EMBL



European Molecular Biology Laboratory

Em 1974 foi criado o EMBL é uma base de dados de sequências nucleotídicas (DNA ou RNA), provenientes de diversas fontes, tais como da literatura científica e de patentes, mas principalmente da submissão direta de sequências pelos pesquisadores ou grupos de seqüenciamento



O **Projeto Genoma Humano (Hugo)** teve início em **1988**, consistiu num esforço internacional para o mapeamento do genoma humano e a identificação de todos os nucleotídeos que o compõem. Após iniciativa dos Institutos Nacionais da Saúde estadunidenses (NIH), centenas de laboratórios de todo o mundo se uniram à tarefa de sequenciar, um a um, os genes que codificam as proteínas do corpo humano e também aquelas sequências de DNA que não são genes. Laboratórios de países em desenvolvimento também participaram do empreendimento com o objetivo de formar mão-de-obra qualificada em genômica.

Science

16 February 2001

Vol. 291 No. 5507
Pages 1145-1434 59

THE HUMAN GENOME



AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

15 February 2001

nature

www.naturejpn.com

the human genome

Nuclear fission

Five-dimensional
energy landscapes

Seafloor spreading

The view from under
the Arctic ice

Career prospects

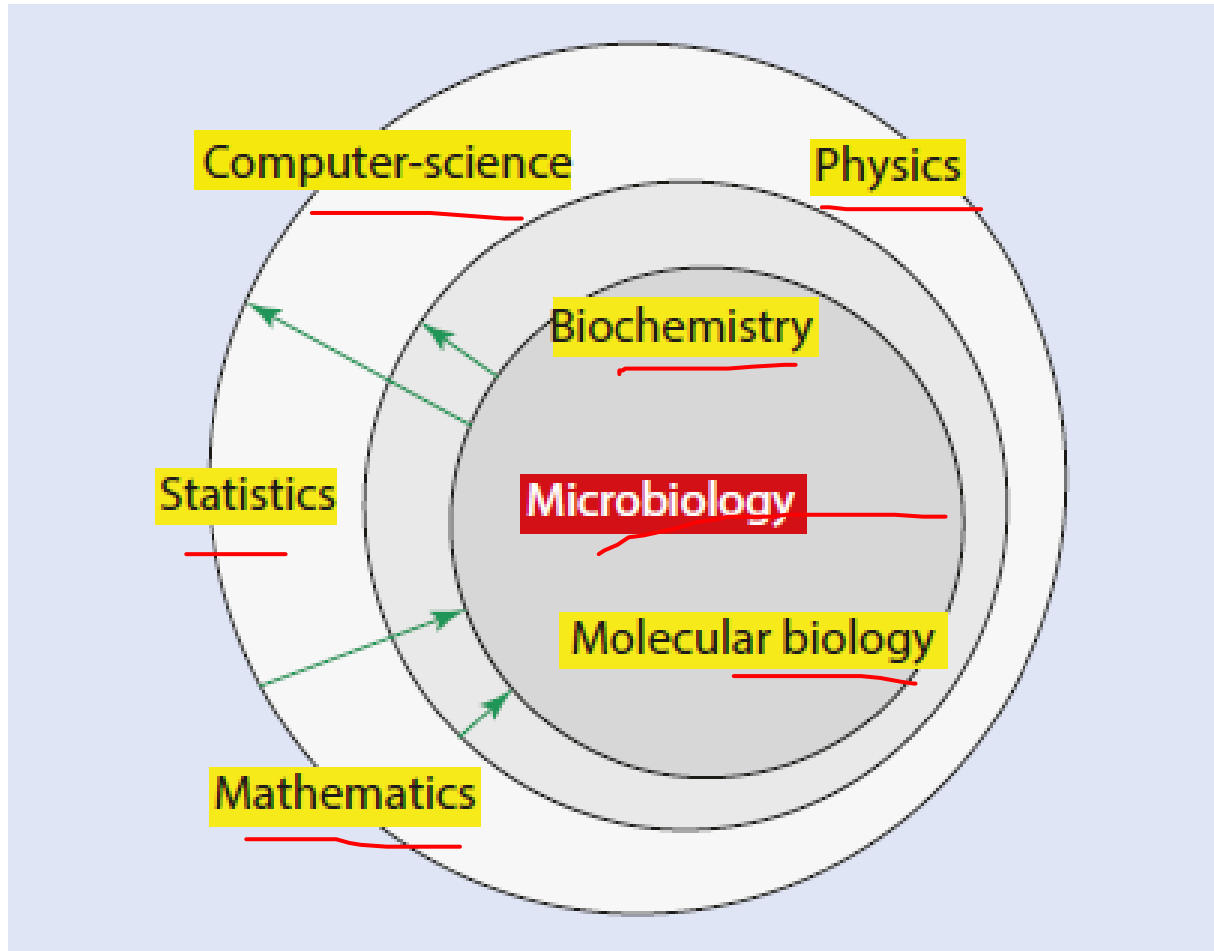
Sequence creates new
opportunities

ヒトゲノム大特集

日米欧などの国際共同チームが人間の全遺伝
情報の概要の解説を11篇の論文にまとめて発表



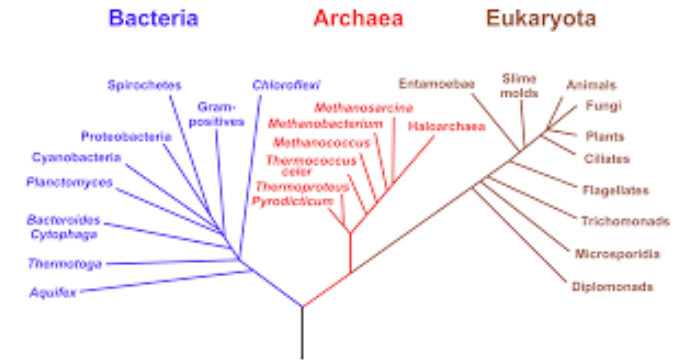
O que é Bioinformática?



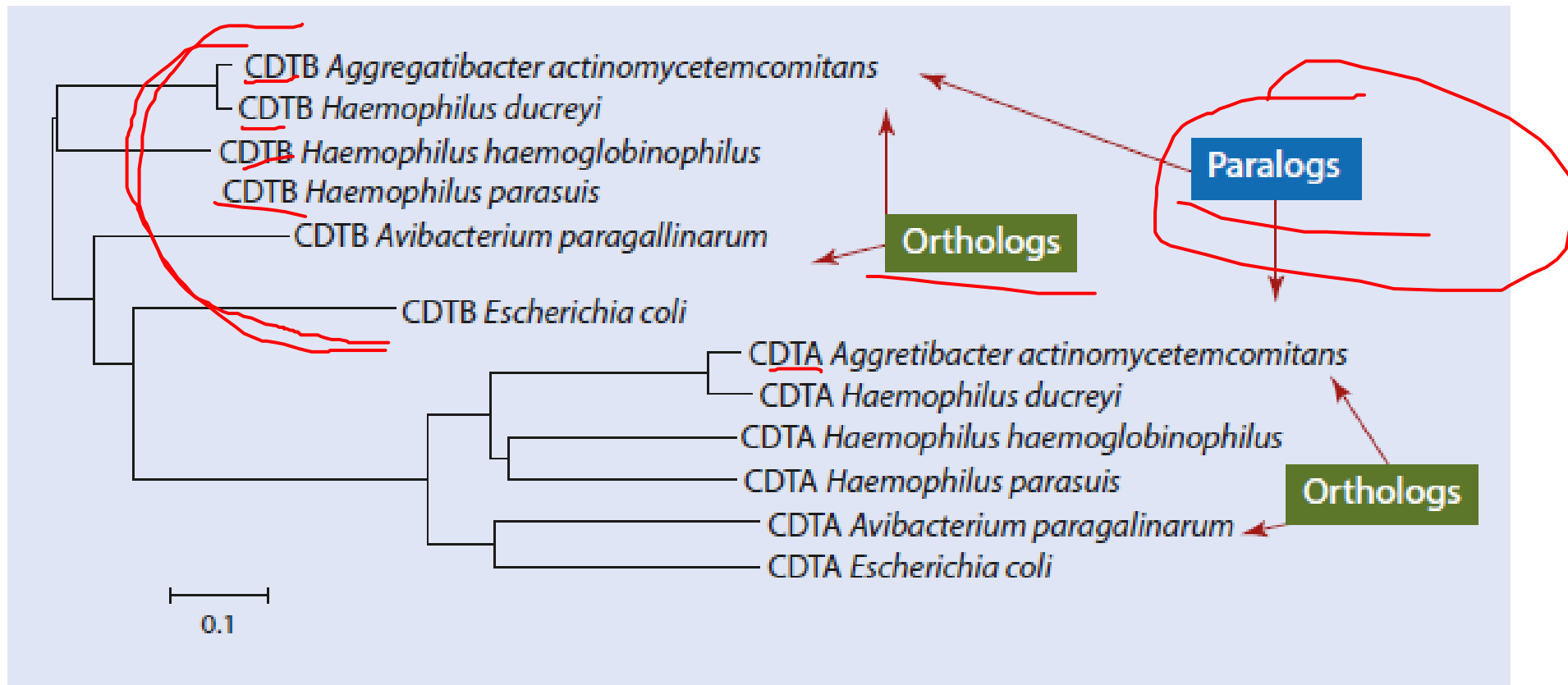
Uma boa comunicação entre as diferentes disciplinas científicas é essencial para os bioinformáticos fazerem pesquisas

- Tem como base o desenvolvimento e aplicação de ferramentas computacionais em analisar todos os tipos de dados biológicos via modelos estatísticos;
- A análise do conteúdo das informações de genomas inteiros. Essas informações incluem os números e os tipos de genes e produtos gênicos, bem como a localização, o número e os tipos de sítios de ligação no DNA e no RNA que possibilitam aos produtos funcionais serem sintetizados na ocasião e no local corretos

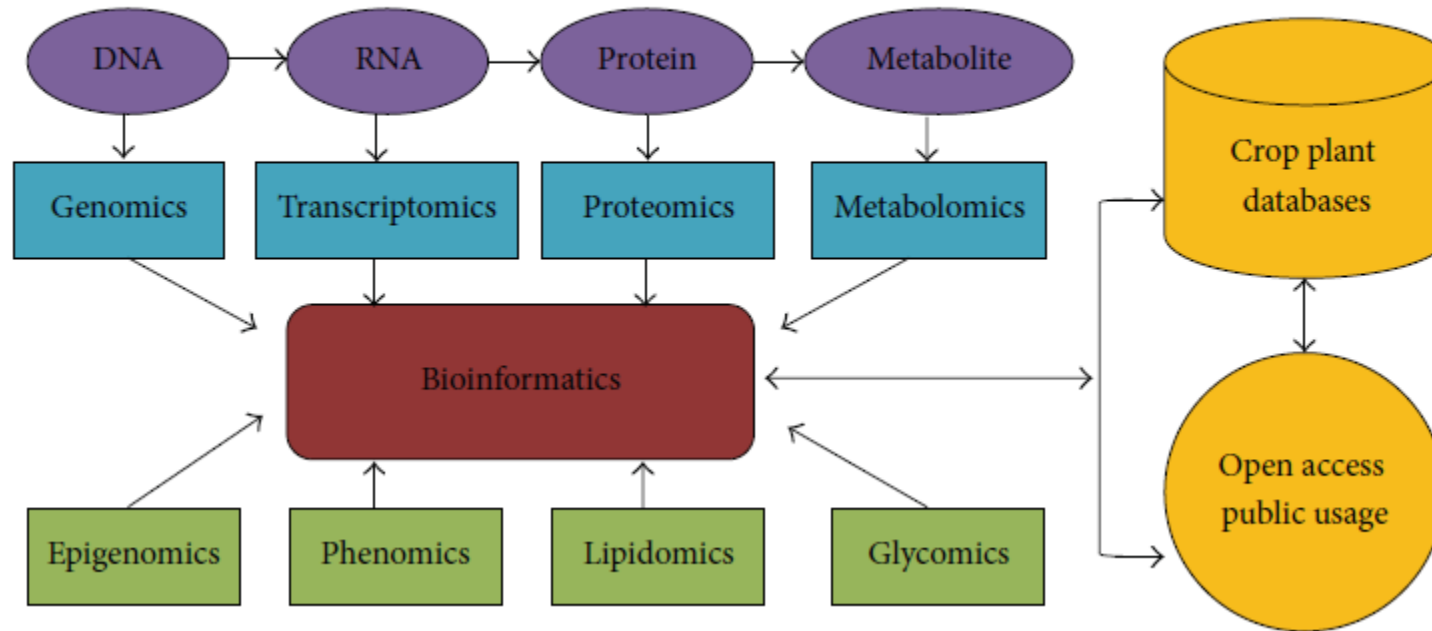
Homologia



- Um conceito biológico chave por trás da bioinformática é a homologia.
- Estruturas homólogas têm o mesmo ancestral ou prevê-se que tenham o mesmo ancestral.
- O conceito de homologia pode ser subdividido em ortologia e paralogia. Esses conceitos foram definidos por Walter M. Fitch e se tornaram os termos-chave usados em bioinformática (Fitch 2000).
- Genes ou proteínas ortólogos são homólogos que divergiram após eventos de especiação, enquanto parálogos são homólogos que divergiram como consequência da duplicação do gene.
- Em procariotos, várias cópias de um gene com a mesma função geralmente não são toleradas e, como consequência, parálogos da mesma espécie terão funções diferentes.



Um exemplo é mostrado na Figura acima da filogenia das proteínas na toxina de distensão citoletal (CDT). O CDT é produzido por alguns patógenos bacterianos, como espécies de *Campylobacter* e *Haemophilus*. CDT não foi encontrado em bactérias Gram-positivas. A CDT pode induzir a parada do ciclo celular G2 / M, fragmentação da cromatina, distensão celular e aumento do núcleo de células eucarióticas. CDT é um heterotrímero que consiste em cadeias de proteínas CDTA, CDTB e CDTC. As três proteínas são homólogas, mas distantemente relacionadas. Pela figura pode-se ver que todas as proteínas CDTA estão relacionadas nas diferentes espécies comparadas e, portanto, são ortólogas. Além disso, as proteínas CDTB estão relacionadas em diferentes espécies, e também são ortólogas de acordo com a definição. No entanto, as proteínas CDTA e CDTB estão mais distantemente relacionadas e são parálogas por estarem presentes na mesma espécie, mas com funções diferentes.



Descrição do fluxograma para integração de plataformas ômicas biológicas com bioinformática, vinculando bancos de dados de plantas agrícolas e uso público de acesso aberto.

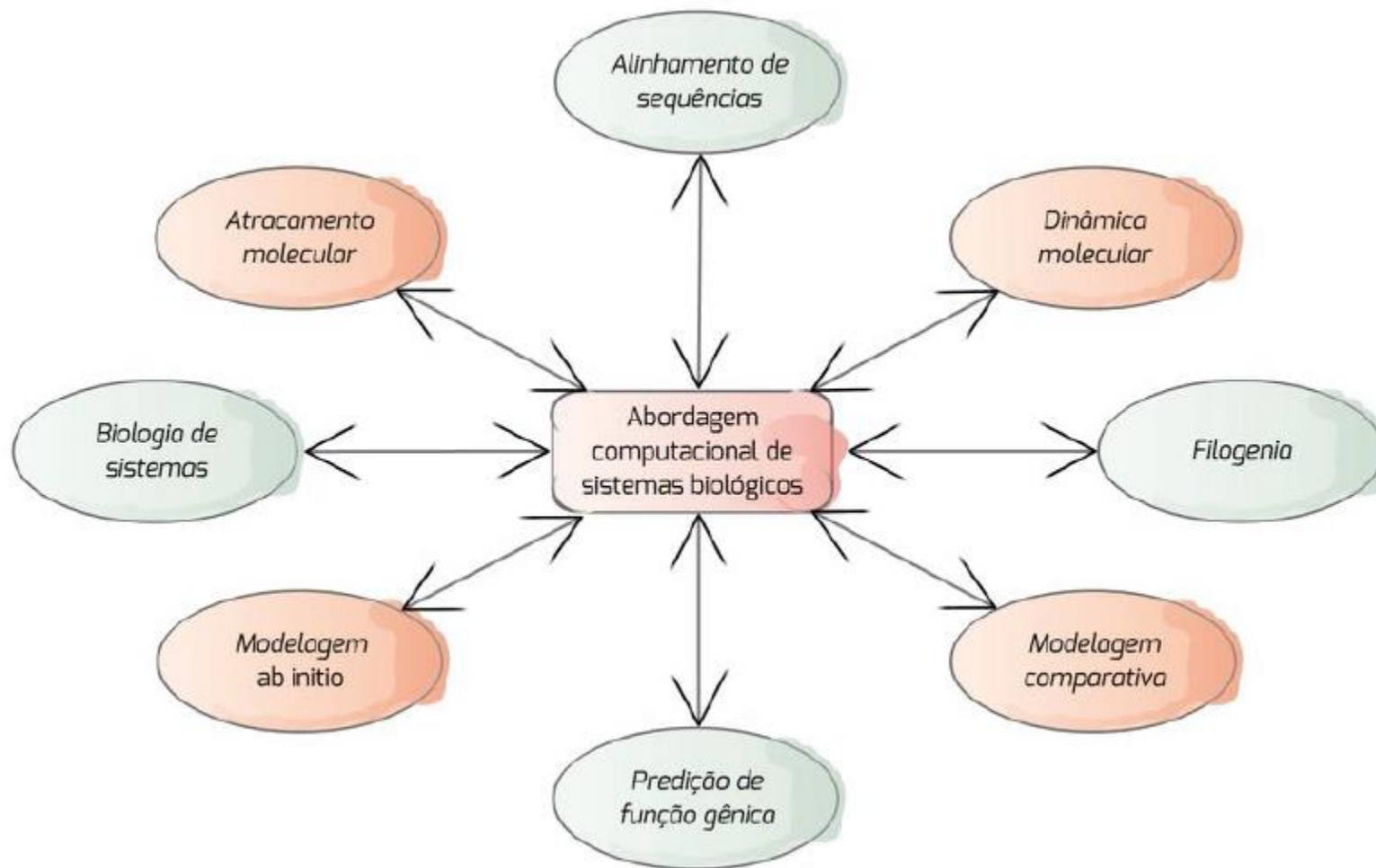
Arun Prabhu Dhanapal and Mahalingam Govindaraj

Hindawi Publishing Corporation

Genetics Research International

Volume 2015, Article ID 684321, 15 pages

<http://dx.doi.org/10.1155/2015/684321>



Representação de algumas das principais áreas da bioinformática. As metodologias que lidam majoritariamente com estruturas 3D estão representadas em laranja, enquanto as metodologias envolvidas principalmente com sequências estão representadas em verde. Devemos lembrar, contudo, que esta separação é imperfeita. Por exemplo, a modelagem comparativa parte de sequências, a função de um gene pode ser determinada pela estrutura da proteína associada.

Key concepts

Algorithm: A logical sequence of instructions needed to execute a task.

Gaps: Regions identified by “-” that represent indels.

Indels: Insertions and deletions of character.

Matches: Corresponding regions between two different sequences.

Mismatches: Regions with non-identical characters in different sequences.

Gap penalty (GP): Parameter needed to assign a score to a gap.

Identity: Percentage of similar characters between two sequences.

Similarity: Degree of resemblance between sequences based on identity.

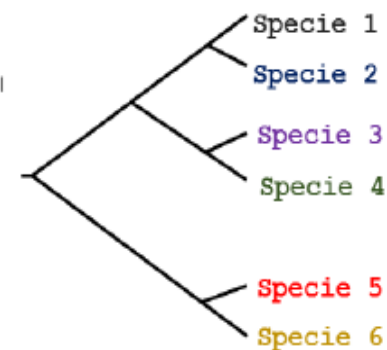
Homology: Evolutionary hypothesis between two sequences that can be derived from a common ancestor.

Paralogs: Genes that diverged by duplication in the genome of the same species.

Orthologs: Genes from a common ancestor that diverged by speciation.

I - Phylogenetic Analysis

```
|1-----10-----20-----30|
>Specie 1| TGCATCTTGCTGGATGCTGCTCTGCTCTCA
>Specie 2| AGCATGTTTCTGGGAGCTGCACTTGTATCT
>Specie 3| AGCATCTTGCTGAAAGCTGCACTTCTTTCT
>Specie 4| ---ATCTTGCTGAAAGCTGCACTTCTTTCT
>Specie 5| TACATGAAGCTGATAGCTGCACTCCTTTCT
>Specie 6| -----TAGCTGATAGCTGCATTTTCATCCT
                ***   *****   *   *
```



II - Identification of genetic variants related to diseases

```
|1-----10-----15|
>Resitant| ATCATCTTTGGTGTT
>Resistant| ATCATCTTTGGTGTT
>Susceptible| ATCATC---GGTGTT
                *****   *****
```

III - Secondary structure prediction and identification of conserved residues

```
|1-----10-----20-----33|
Rat | --YPTFHGPISRVRAAQLVQLQGPDAHCVFLVR
Mouse| --SATYMGPVTRQEAQTRLQGR---HGMFLVR
Human| ----YYGKVTIRHQAEMALNERGH--EGDFLI
                * * * : * ** *
```

Figure 2. Sequence alignment and some of its applications. Figure modified from Junqueira et al. (2014).

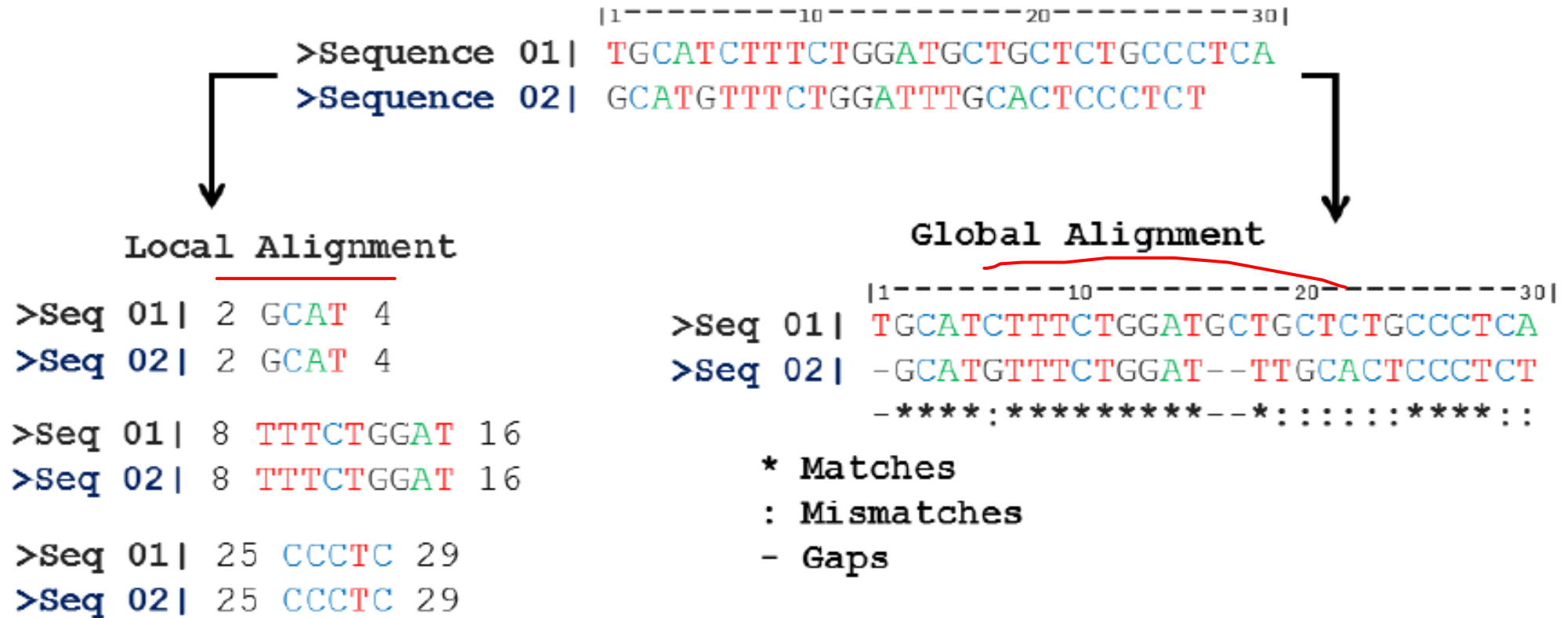


Figure 3. Global and local alignment of amino acid sequences. Figure modified from Prosdocimi et al. (2002).

Alinhamento global busca o melhor alinhamento em toda a extensão das duas sequências comparadas.

Alinhamento local busca somente alinhamento de regiões de alta similaridade, não importando as sequências adjacentes a estas regiões

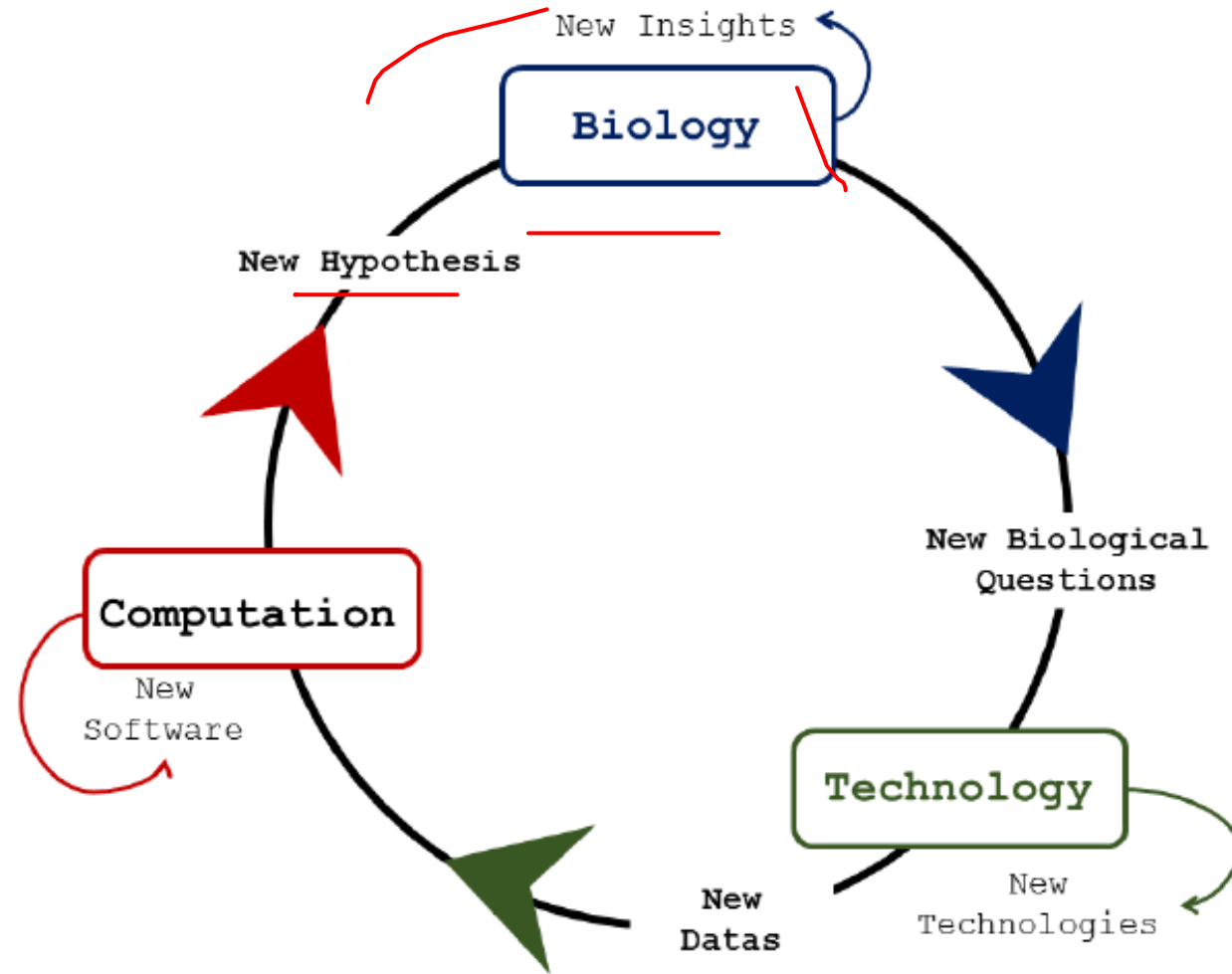


Figure 15. Systems biology as a multidisciplinary science. Figure modified from Institute for Systems Biology (2016).

Table 2.1 Selection of DNA sequencing strategies and methods in relation to capacity and cost

Problem	Method	Read length (nt)	Capacity	Cost (related to capacity)
1. Single genes, sequencing of BAC, FOS ^a or plasmid inserts (cloning)	Sanger sequencing of PCR amplicons	500–1000	Low	High
2. Genomes and 16S metagenomics	Illumina	150–600	High	Low
	Ion torrent	400–600	High	Low
3. Single-cell sequencing	Single-cell dissection, DNA extraction, phi29 amplification, and Illumina sequencing	150–251	High	High
4. Closing of single genomes	PacBio® (Pacific Biosciences)	15,000–100,000	Medium	High
	Nanopore	1000.000	Medium	High

Diferente plataformas utilizadas para o sequenciamento do DNA

TABLE 1: List of crop plant genomes available to public with URL information and references.

Name of crop plant	Consortium/initiative	URL	References
Alfalfa (<i>Medicago sativa</i>)	Consortium	http://www.alfalfa-genome.org/www/	Young et al., 2011 [7]
Apple (<i>Malus domestica</i>)	Consortium	http://www.rosaceae.org/species/malus/malus_spp	Velasco et al., 2010 [8]
Banana (<i>Musa acuminata</i>)	The Global Musa Genomics Consortium	http://www.musagenomics.org/	D'Hont et al., [9]
Barley (<i>Hordeum vulgare</i>)	International Barley Genome Sequencing Consortium	http://www.public.iastate.edu/~imagefpc/IBSC%20Webpage/IBSC%20Template-home.html	The International Barley Genome Sequencing Consortium [10]
Cacao (<i>Theobroma cacao</i>)	Consortium	http://www.cacaogenomedb.org/	Argout et al., 2011 [11]
Cannabis (<i>Cannabis sativa</i>)	Consortium	http://genome.ccb.utoronto.ca/index.html?org=C.+sativa&db=canSat3&hgsid=11252	van Bakel et al., 2011 [12]
Castor bean (<i>Ricinus communis</i>)	TIGR	http://castorbean.jcvi.org/	Chan et al., 2010 [13]
Chickpea (<i>Cicer arietinum</i>)	Consortium (ICRISAT-BGI)	http://www.icrisat.org/gt-bt/ICGGC/GenomeManuscript.htm	Varshney et al., 2013 [14]
Chocolate (<i>Theobroma cacao</i>)	Consortium	http://www.cacaogenomedb.org/main	Argout et al., 2011 [11]
Cotton (<i>Gossypium raimondii</i>)	BGI	http://www.cottondb.org/wwwroot/cdbhome.php	Wang et al., 2012 [15]
Common bean (<i>Phaseolus vulgaris</i> L.)	Consortium	http://www.phytozome.net/commonbean.php	Schmutz et al., 2014 [16]
Crucifer (<i>Thellungiella parvula</i>)	Consortium	http://www.brassica.info/info/events.php	Dassanayake et al., 2011 [17]
Cucumber (<i>Cucumis sativus</i>)	International Cucurbit Genomics Initiative (ICuGI)	http://www.icugi.org/	Huang et al., 2009 [18]
Date palm (<i>Phoenix dactylifera</i>)	Consortium	http://qatar-weill.cornell.edu/research/datepalmGenome/download.html	Al-Dous et al., 2011 [19]
Wheat (<i>Triticum aestivum</i>)	International Wheat Genome Sequencing Consortium (IWGSC)	http://www.wheatgenome.org/	International Wheat Genome Sequencing Consortium, 2014 [20]
Foxtail millet (<i>Setaria italica</i>)	Beijing Genomics Institute and the Joint Genomes Institute	http://www.phytozome.net/foxtailmillet.php	Zhang et al., [21] Bennetzen et al. [22]
Grape (<i>Vitis vinifera</i>)	Consortium	http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/	Jaillon et al., 2007 [23]
<i>Jatropha curcas</i> L.	Consortium	http://www.kazusa.or.jp/jatropha/	Sato et al., 2011 [24]
Lotus (<i>Lotus japonicas</i>)	Consortium	http://www.kazusa.or.jp/lotus/	Sato et al., 2008 [25]
Maize (<i>Zea mays</i>)	Consortium	http://www.maizegdb.org/	Schnable et al., 2009 [26]
Model Grass (<i>Brachypodium distachyon</i>)	Consortium	http://www.brachypodium.org/	International Brachypodium Initiative, 2010 [27]
Mosses (<i>Physcomitrella patens</i>)	JGI	http://genome.jgi-psf.org/PhypaL1/PhypaL1.home.html	Rensing et al., 2008 [28]
Mouse ear cress (<i>Arabidopsis thaliana</i> , <i>Arabidopsis lyrata</i>)	The Arabidopsis Genome Initiative (2000)	http://www.arabidopsis.org/	The Arabidopsis Genome Initiative, 2000 [29]; Cao et al., 2011 [30]; Hu et al., 2011 [31]
Papaya (<i>Carica papaya</i>)	Consortium	http://www.plantgdb.org/CpGDB/	Ming et al., 2008 [32]
Peach (<i>Prunus persica</i>)	International Peach Genome Initiative	http://www.rosaceae.org/peach/genome	International Peach genome initiative 2013 [33]

TABLE 1: Continued.

Name of crop plant	Consortium/initiative	URL	References
Pigeon pea (<i>Cajanus cajan</i>)	International Initiative for Pigeonpea Genomics (IIPG)	http://www.icrisat.org/gt-bt/iipg/Home.html	Varshney et al., 2011 [34]
Poplar (<i>Populus trichocarpa</i>)	JGI	http://genome.jgi-psf.org/Poptr1.1/Poptr1.1.home.html	Tuskan et al., 2006 [35]
Potato (<i>Solanum tuberosum</i>)	Consortium (PGSC)	http://www.potatogenome.net/index.php/Main_Page	The Potato Genome Sequencing Consortium, 2011 [36]
Rape seed (<i>Brassica napus</i>)	Consortium (MGBP)	http://www.plantgdb.org/BrGDB/	Wang et al., 2011 [37]
Rice (<i>Oryza sativa</i> ssp. <i>indica</i> and <i>japonica</i>)	Consortium (IRGSP)	http://rgp.dna.affrc.go.jp/E/IRGSP/index.html	Yu et al., 2002 [38]; Goff et al., 2002 [39]
Sorghum (<i>Sorghum bicolor</i>)	JGI	http://www.plantgdb.org/SbGDB/	Paterson et al., 2009 [40]
Soybeans (<i>Glycine max</i>)	JGI	http://www.phytozome.net/soybean_er.php	Schmutz et al., 2010 [41]
Strawberry (<i>Fragaria vesca</i>)	Consortium	http://www.strawberrygenome.org/	Shulaev et al., 2011 [42]
<u>Tomato</u> (<u><i>Solanum lycopersicum</i></u>)	<u>Consortium</u>	<u>http://solgenomics.net/organism/Solanum_lycopersicum/genome</u>	The <u>Tomato Genome Consortium</u> 2012 [43]
Watermelon (<i>Citrullus lanatus</i>)	International Watermelon Genomics Initiative	http://www.iwgi.org/	Guo et al., 2013 [44]
Mung bean (<i>Vigna radiata</i>)	Not available	http://vigra.comparative-legumes.org/	Kang et al., 2014 [45]

TABLE 2: List of databases and their information of application to crop plant research community.

Name of databases	Application
AgBase—a unified resource for functional analysis in agriculture	Search and analyze functional genomics datasets in agricultural species
AutoSNPdb—an annotated single nucleotide polymorphism database for crop plants	Identify SNPs from assembled EST sequences for the crops rice, barley, and <i>Brassica</i>
BarleyBase—an expression profiling database for plant genomics	Analyze and visualize plant microarray data
BBGD—an online database for blueberry genomic data.	It stores both EST and microarray data and allows scientists to correlate expression profiles with gene function
BIOGEN BASE—CASSAVA	A web accessible resource for investigating cassava phenomics and genomics information
CastorDB—a comprehensive knowledge base for <i>Ricinus communis</i> .	CastorDB provides a user friendly comprehensive resource on castor with particular emphasis on its genome, transcriptome, and proteome and on protein domains, pathways, protein localization, presence of sumoylation sites, expression data, and protein interacting partners
ChromDB—The Chromatin Database	Locate chromatin-associated proteins, including RNAi-associated proteins, for a broad range of organisms
CR-EST—a resource for crop ESTs	Search for sequence, classification, clustering, and annotation data of crop EST projects
CSRDB—a small RNA integrated database and browser resource for cereals	Search for sequence information on rice, maize, and other cereal crops small RNAs
DEBDOM—Database Exploring Banana Diversity of Manipur	The database DEBDOM provides a sophisticated web base access to the details of the taxonomy, morphological characteristics, utility, and sites of collection of <i>Musa</i> genotypes
DRASTIC—Database Resource for the Analysis of Signal Transduction in Cells	Search for information of plant gene expression in response to pathogens and environmental changes
FLAGdb++—A Database for the Functional Analysis of the Arabidopsis Genome	Search and visualize data for high-throughput functional analysis of <i>Arabidopsis</i> , rice, and other plant genomes
GabiPD—a plant integrative “omics” database	Search for comprehensive and extensive information on various plant genomes generated by a German collaborative network of plant genomics research
GCP—The Generation Challenge Programme	An online resource documenting stress-responsive genes comparatively across plant species
GDR—Genome Database for Rosaceae	A central repository of curated and integrated genetics and genomics data of Rosaceae, which includes apple, cherry, peach, pear, raspberry, rose, and strawberry
GeneCAT—gene co-expression analysis toolbox	Novel web tools that combine BLAST and coexpression analyses
GeneSeqer@PlantGDB—gene structure prediction in plant genomes	Predict gene structures of plant genomes
GERMINATE	A generic database for integrating genotypic and phenotypic information for plant genetic resource collections
GGT—Graphical GenoTypes	Software for visualization and analysis of genetic data
GrainGenes—The genome database for small-grain crops	Search for molecular and phenotypic information on wheat, barley, rye, triticale, and oats
Gramene—a resource for comparative grass genomics	Curated resource for genetic, genomic, and comparative genomics data for the major crop species, including rice, maize, wheat, and many other plant (mainly grass) species
MaizeGDB—the Community Database for Maize Genetics and Genomics	Search genetic and genomic information about maize

TABLE 2: Continued.

Name of databases	Application
MANET—The Molecular Ancestry Network	Tracing evolution of protein architecture in metabolic networks
Medicago—A database for personalized data mining of the model legume <i>Medicago truncatula</i> transcriptome	Search for integrated genomic, genetic, and biological information on cool season legume <i>Medicago truncatula</i> (Mt)
MetaCrop—a detailed database of crop plant metabolism	A database that summarizes diverse information about metabolic pathways in crop plants and allows automatic export of information for the creation of detailed metabolic models
MetaCrop 2.0—managing and exploring information about crop plant metabolism.	It contains information about seven major crop plants with high agronomical importance and two model plants; MetaCrop is intended to support research aimed at the improvement of crops for both nutrition and industrial use
Narcisse—a mirror view of conserved syntenies	A database dedicated to the study of genome conservation
NIASGBdb—National Institute of Agrobiological Sciences Gene Bank DataBase	Find information about agricultural plant genetics and diseases
P3DB—Plant Protein Phosphorylation Database	Find information about protein phosphorylation in plants
Panzea—a database and resource for molecular and functional diversity in the maize genome	Search for information on relationship between genotype and functional phenotype variations
Pepper EST database—in silico exploitation of EST data to extensively score genes of <i>Capsicum annuum</i>	Comprehensive in silico tool for analyzing the chili pepper (<i>Capsicum annuum</i>) transcriptome
PIP—a database of potential intron polymorphism markers	A database of potential intron polymorphism markers in plants
PLACE—plant cis-acting regulatory DNA elements	Search for documented motifs found in plant cis-acting regulatory DNA elements
Plant snoRNA database	Search for comprehensive information on small nucleolar RNAs in plants
PlantCARE—a database of plant cis-acting elements	Search for information on plant cis-acting regulatory elements, transcription sites, enhancers, and repressors
PlantTFDB—Plant Transcription Factor Databases	A comprehensive plant transcription factor database
PlantTribes—a gene and gene family resource for comparative genomics in plants	A plant gene family database based on the inferred proteomes of five sequenced plant species: <i>Arabidopsis thaliana</i> , <i>Carica papaya</i> , <i>Medicago truncatula</i> , <i>Oryza sativa</i> , and <i>Populus trichocarpa</i>
PLecDom—Plant Lectin Domains server	Find information about plant lectin domains
PlnTFDB—Plant Transcription Factor Database	Find information about transcription factors in plants
PmiRKB—Plant MicroRNA Knowledge Base	Find information about plant microRNAs
PMRD—Plant MicroRNA Database	Find information about microRNA sequences and targets in plants
PODB—the Plant Organelles Database	Search a collection of visualized plant organelles and protocols for plant organelle research
POGs/PlantRBP—a resource for comparative genomics in plants	Search for information on putative orthologous proteins among rice, maize, and <i>Arabidopsis</i> with emphasis on RNA-binding proteins
PoMaMo—a comprehensive database for potato genome data	Search for comprehensive genomic information on potato
PREP Suite—Predictive RNA Editor for Plants	Use to predict sites of RNA editing in plants
PRGDB—Plant Resistance Genes DataBase	Find information about genes involved in plant defense mechanisms
pssRNAMiner—a plant short small RNA regulatory cascade analysis server	Identify both the clusters of phased small RNAs as well as the potential phase-initiator
RadishBase—a database for genomics and genetics of radish.	A database containing radish pathways predicted from unigene sequences is also included in RadishBase
RoBuST—an integrated genomics resource for the root and bulb crop families Apiaceae and Alliaceae.	The RoBuST database has been developed to initiate a platform for collecting and organizing genomic information useful for RBV (root and bulb vegetables) researchers

1. Criado em 4 de novembro de 1998, como uma divisão da NLM e NIH

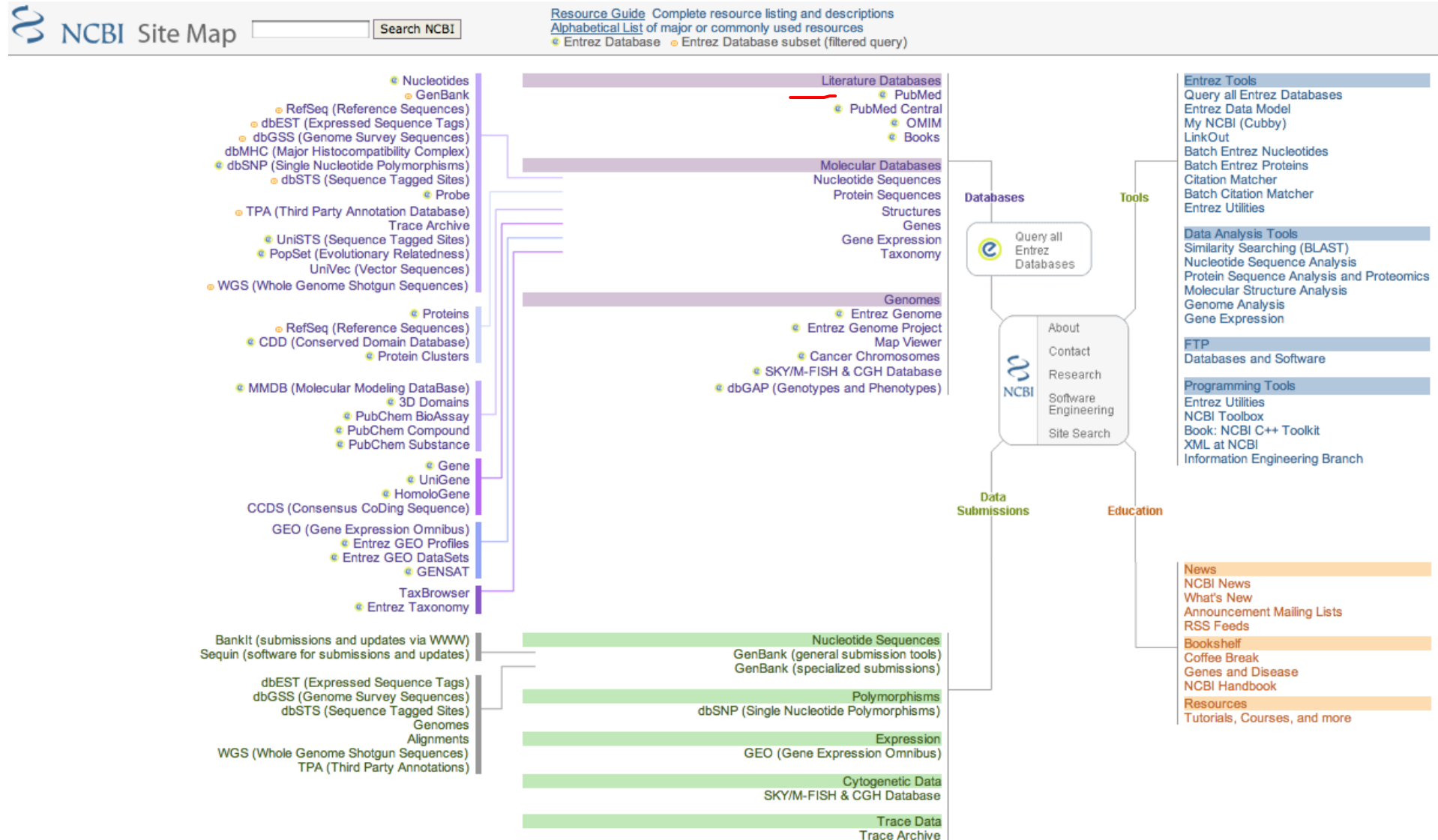
O que faz:

- I. Pesquisa a nível molecular em ciências biomédicas, empregando métodos matemáticos e computacionais.
- II. Colaboração com outros institutos do NIH, com a academia, indústria e outras agências
Comunicação científica, por meio de reuniões, workshops e conferências.
- III. Formação em pesquisa básica e aplicada em biologia computacional. Desenvolve, distribui, apóia e coordena o acesso a muitos bancos de dados e software para a comunidade científica e médica.
- IV. Desenvolve e promove padrões para intercâmbio de dados.

Manual NCBI: <http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=handbook>

<http://www.ncbi.nlm.nih.gov/Sitemap>

NCBI: Site map

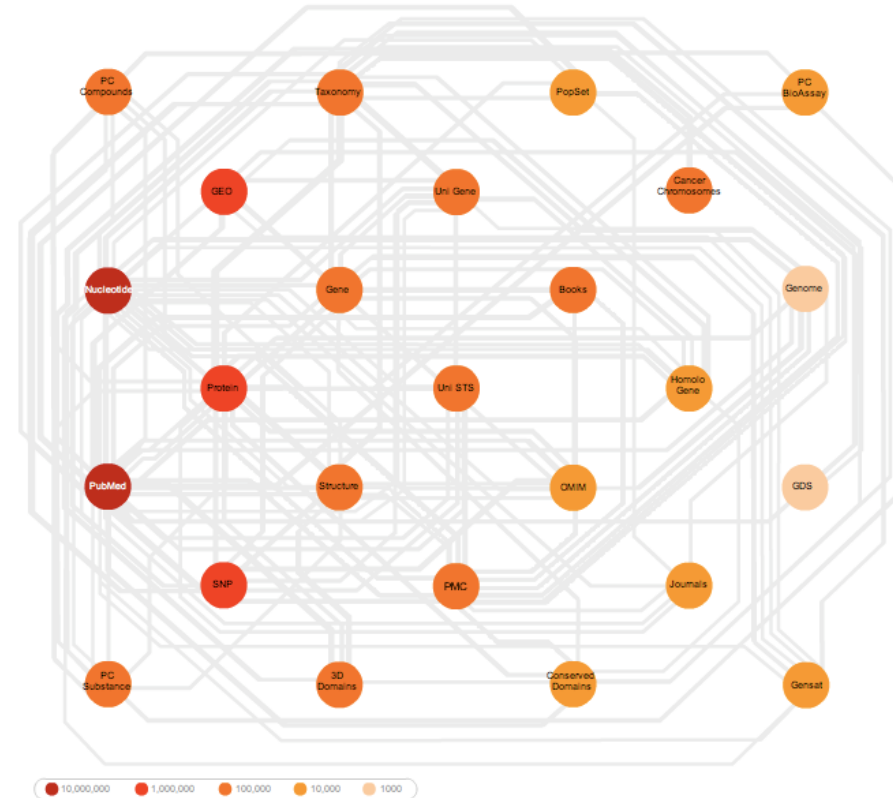


NCBI: Entrez The life science search engine

- Not a database itself, but a interface to query several databases at once:
 - Literature
 - ADN

 - Proteins

 - 3D structures
 - Protein domains
 - Population studies
 - Gene expression data
 - Genome data
 - Taxonomic information



BLAST

<http://blast.ncbi.nlm.nih.gov/>

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Arabidopsis thaliana](#)
- [Oryza sativa](#)
- [Bos taurus](#)
- [Danio rerio](#)
- [Drosophila melanogaster](#)
- [Gallus gallu](#)
- [Pan troglod](#)
- [Microbes](#)
- [Apis mellife](#)

Basic BLAST

Choose a BLAST program to run.

[nucleotide blast](#)

Search a **nucleotide** database using a **nucleotide** query
Algorithms: blastn, megablast, discontinuous megablast

[protein blast](#)

Search **protein** database using a **protein** query
Algorithms: [blastp](#), [psi-blast](#), [phi-blast](#)

[blastx](#)

Search **protein** database using a **translated nucleotide** query

[tblastn](#)

Search **translated nucleotide** database using a **protein** query

[tblastx](#)

Search **translated nucleotide** database using a **translated nucleotide** query

<http://blast.ncbi.nlm.nih.gov/>

BLAST

- Ferramenta de busca de similaridade de sequências.
- Consulta de sequências em Banco de dados biológicos

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

- | | | |
|---|--|---|
| <input type="checkbox"/> Human | <input type="checkbox"/> Oryza sativa | <input type="checkbox"/> Gallus gallu |
| <input type="checkbox"/> Mouse | <input type="checkbox"/> Bos taurus | <input type="checkbox"/> Pan troglod |
| <input type="checkbox"/> Rat | <input type="checkbox"/> Danio rerio | <input type="checkbox"/> Microbes |
| <input type="checkbox"/> Arabidopsis thaliana | <input type="checkbox"/> Drosophila melanogaster | <input type="checkbox"/> Apis mellife |

Basic BLAST

Choose a BLAST program to run.

nucleotide blast	Search a nucleotide database using a nucleotide query <i>Algorithms: blastn, megablast, discontinuous megablast</i>
protein blast	Search protein database using a protein query <i>Algorithms: blastp, psi-blast, phi-blast</i>
blastx	Search protein database using a translated nucleotide query
tblastn	Search translated nucleotide database using a protein query
tblastx	Search translated nucleotide database using a translated nucleotide query



► [NCBI/BLAST/blastp suite: BLASTP programs search protein databases using a protein query.](#) [more...](#)

[Reset page](#)

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

[Clear](#)

NP_006735

Query subrange

From

To

Or, upload file

[Seleccionar archivo](#) ningún ar...ccionado

Job Title

NP_006735:retinol-binding protein 4, plasma...

Enter a descriptive title for your BLAST search

Choose Search Set

Database

Non-redundant protein sequences (nr)

Organism
Optional

Enter organism name or id—completions will be suggested

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Entrez Query
Optional

"Homo sapiens"

Enter an Entrez query to limit search

Program Selection

Algorithm

- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm

BLAST

Search database nr using Blastp (protein-protein BLAST)

Show results in a new window

► [Algorithm parameters](#)

Online exercise (Hint: use the option for advanced search)

1. Go to <http://www.ncbi.nlm.nih.gov/>
2. Search for Hemoglobin Subunit Beta (HBB) mRNA in humans
3. Select nucleotide from the result screen
4. Show the entry NM_000518.4
5. Identify the different pieces of information from the results screen.

Anotação

- Postular função para produto de gene;
- Predizer estruturas do genoma e suas funções.

Anotam-se:

- Genes que codificam proteínas
- tRNAs
- rRNAs
- ORFs hipotéticos
- Clusters de GC
- Repetições
- Codon usage
- Promotores

GenBank

- Banco de dados de sequência de nucleotídeos
- Parte da "Colaboração Internacional de Banco de Dados de Sequência de Nucleotídeos", que também é integrada por EMBL y el DDBJ
- Em sua versão mais recente (15 de abril de 2016), existem 211.423.912.047 bases, a partir de 193.739.511 sequências relatadas.
- É atualizado aprox. a cada 2 meses
- As sequências são enviadas diretamente por seus próprios autores.
- Há uma verificação de qualidade mínima realizada pela equipe do GenBank, e todas as informações associadas a uma sequência dependem totalmente de seus remetentes
- O nível de redundância é alto

Record format for a GenBank entry

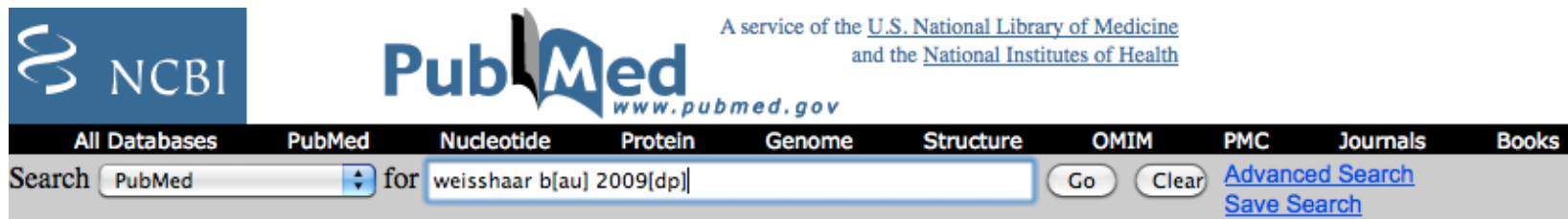
```
LOCUS SCU49845 5028 bp DNA PLN 21-JUN-1999
DEFINITION Saccharomyces cerevisiae TCP1-beta gene, partial cds, and Axl2p
            (AXL2) and Rev7p (REV7) genes, complete cds.
ACCESSION U49845
VERSION U49845.1 GI:1293613
KEYWORDS .
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
          Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
          Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 5028)
AUTHORS Torpey,L.E., Gibbs,P.E., Nelson,J. and Lawrence,C.W.
TITLE Cloning and sequence of REV7, a gene whose function is required for
      DNA damage-induced mutagenesis in Saccharomyces cerevisiae
JOURNAL Yeast 10 (11), 1503-1509 (1994)
MEDLINE 95176709
PUBMED 7871890
REFERENCE 2 (bases 1 to 5028)
AUTHORS Roemer,T., Madden,K., Chang,J. and Snyder,M.
TITLE Selection of axial growth sites in yeast requires Axl2p, a novel
      plasma membrane glycoprotein
JOURNAL Genes Dev. 10 (7), 777-793 (1996)
MEDLINE 96194260
PUBMED 8846915
REFERENCE 3 (bases 1 to 5028)
AUTHORS Roemer,T.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-1996) Terry Roemer, Biology, Yale University, New
      Haven CT, USA
FEATURES             Location/Qualifiers
     source            1..5028
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="IX"
                     /map="9"
     CDS                <1..206
                     /codon_start=3
                     /product="TCP1-beta"
                     /protein_id="AAA98665.1"
                     /db_xref="GI:1293614"
                     /translation="SSIYNGISTSGLDLNNGTIADMRQLGIVESYKLRKRAVSSASEA
                     AEVLLRVDNIIIRARPRTANRQHM"
     gene                687..3158
                     /gene="AXL2"
     CDS                687..3158
                     /gene="AXL2"
                     /note="plasma membrane glycoprotein"
                     /codon_start=1
                     /function="required for axial budding pattern of S.
BASE COUNT            1510 a 1074 c 835 g 1609 t
ORIGIN
1 gatcctccat atacaacggt atctccaact cagggtttaga tctcaacaac ggaaccattg
61 cggacatgag acagttaggt atcgtcgaga gttacaagct aaaacgagca gtagtacgct
121 ctgcatgga agccctgaa gttctactaa ggggtgataa cttcatccgt gcaagaccaa
```

<http://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html>

PubMed



<http://www.pubmed.gov/>



PubMed

PubMed Search
Create RSS Create alert Advanced Help

Summary ▾ 20 per page ▾ Sort by Most Recent ▾

Send to: ▾ Filters: [Manage Filters](#)

Search results

Items: 1 to 20 of 23

<< First < Prev Page 1 of 2 Next > Last >>

1. [Analysis of Sensitive CO2 Pathways and Genes Related to Carbon Uptake and Accumulation in Chlamydomonas reinhardtii through Genomic Scale Modeling and Experimental Validation.](#)

Winck FV, Melo DO, Riaño-Pachón DM, Martins MC, Caldana C, Barrios AF.
Front Plant Sci. 2016 Feb 9;7:43. doi: 10.3389/fpls.2016.00043. eCollection 2016.
PMID: 26904035 [Free PMC Article](#)
[Similar articles](#)

2. [EEF1D modulates proliferation and epithelial-mesenchymal transition in oral squamous cell carcinoma.](#)

Flores IL, Kawahara R, Miguel MC, Granato DC, Domingues RR, Macedo CC, Carnielli CM, Yokoo S, Rodrigues PC, Monteiro BV, Oliveira CE, Salmon CR, Nociti FH Jr, Lopes MA, Santos-Silva A, **Winck FV**, Coletta RD, Paes Leme AF.
Clin Sci (Lond). 2016 May 1;130(10):785-99. doi: 10.1042/CS20150646. Epub 2016 Jan 28.

Find related data

Database:

[Find items](#)

Search details

"winck fv"[au]

[Search](#)


[See more...](#)

Recent Activity

[Turn Off](#) [Clear](#)

Q "winck fv"[au] (23)

PubMed

 PubMed [Advanced](#) [Help](#)

[Abstract](#) Send to:

[Brief Funct Genomic Proteomic.](#) 2008 Jul;7(4):312-21. doi: 10.1093/bfgp/eln023. Epub 2008 May 29.

The untiring search for the most complete proteome representation: reviewing the methods.




[Martins de Souza D¹](#), [Oliveira BM](#), [Castro-Dias E](#), [Winck FV](#), [Horiuchi RS](#), [Baldasso PA](#), [Caetano HT](#), [Pires NK](#), [Marangoni S](#), [Novello JC](#).

[+ Author information](#)

Abstract

Proteomic research has proved valuable for understanding the molecular mechanisms of biological processes, as well as in the search for biomarkers for a variety of diseases which lack a molecular diagnostic. While several new approaches are being developed, two-dimensional (2-DE) gel electrophoresis is still one of the most commonly used techniques, despite its many limitations. However, for biomarker research, 2-DE gel electrophoresis alone does not fulfill the necessary pre-requisites. If such a technique is utilized exclusively, a great part of a given proteome remains unseen. Therefore, very precise and sensitive techniques are needed. Here, we present a brief review of known methodologies that try to overcome the limitations of conventional proteome analysis as well as their respective advantages and limitations.

PMID: 18511486 [PubMed - indexed for MEDLINE] [Free full text](#)

Full text links

[FREE FINAL TEXT](#)
OXFORD JOURNALS

Save items

Similar articles

[Review](#) Applications and current challenges of proteom [Amino Acids. 2007]

All about DIGE: quantification technology for different [Expert Rev Proteomics. 2004]

[Review](#) Design of proteome-based studies in combination [Proteomics. 2002]

[Review](#) Clinical proteomics in cancer research-promises [Curr Med Chem. 2008]

Some other databases @ NCBI

- **OMIM:** Online Mendelian Inheritance in Man. Catalogue of human genes and genetic disorders
- **Books:** NCBI offers several dozens of book online
- **Taxonomy:** Taxonomy search engine for the mayor taxonomy divisions.
- **Structure:** Molecular Modelling Database. Contains 3D protein structures.

Entrez



Entrez, The Life Sciences Search Engine.



- Go to the site:

<http://www.ncbi.nlm.nih.gov/>

- Select the database: PubMed

The screenshot shows the top navigation bar of the NCBI website. The NCBI logo is on the left. The main title is "National Center for Biotechnology Information" with links to "National Library of Medicine" and "National Institutes of Health". Below this is a dark blue navigation bar with links for "PubMed", "All Databases", "BLAST", "OMIM", "Books", "TaxBrowser", and "Structure". A search bar is present with a dropdown menu currently set to "All Databases". The dropdown menu is open, showing options: "All Databases", "NCBI Web Site", "PubMed", "Protein", and "Nucleotide". Below the search bar, there are links for "What does NCBI do?" and "Hot Spots".

Entrez

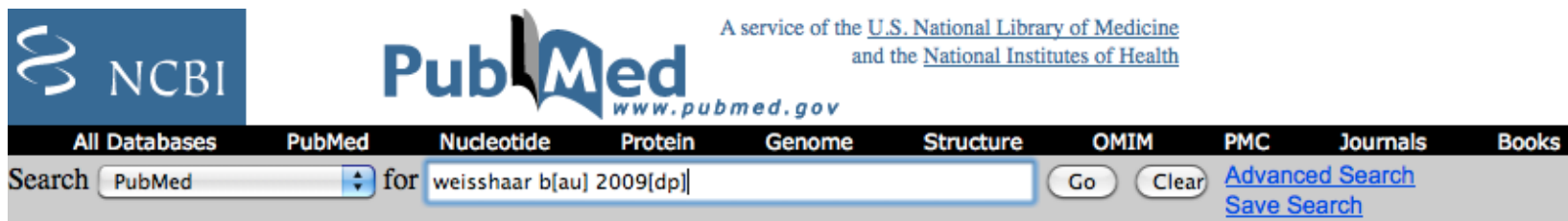


Entrez, The Life Sciences Search Engine.



- Search for the papers of a single author (Weisshaar B) from one specific year (2009). Use the modifiers DP y AU. As it is shown below:

<http://www.ncbi.nlm.nih.gov/>



As an alternative option you can use the “Advance search” option

http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=helppubmed&part=pubmedhelp#pubmedhelp.Search_Field_Descrip

Entrez



Entrez, The Life Sciences Search Engine



- Select the paper shown in the figure:

[Arabidopsis thaliana expresses a second functional flavonol synthase.](#)

Preuss A, Stracke R, **Weisshaar B**, Hillebrecht A, Matern U, Martens S.

FEBS Lett. **2009** Jun 18;583(12):1981-6. doi: 10.1016/j.febslet.2009.05.006. Epub **2009** May 10.

PMID: 19433090 **Free Article**

[Similar articles](#)



You can follow this link to find other papers on the same matter.

- Click on the title of the paper.

Entrez



Entrez, The Life Sciences Search Engine



Abstract ▾

FEBS Lett. 2009 Jun 18;583(12):1981-6. doi: 10.1016/j.febslet.2009.05.006. Epub 2009 May 10.

Arabidopsis thaliana expr

Preuss A¹, Stracke R, Weisshaar B, H

Author information

Abstract

Arabidopsis thaliana L. produces flavonoid dihydroflavonol substrates. A small functional activity was attributed only to mutants still accumulate significant proteins were therefore examined for This enzyme is therefore likely responsible and biochemical data demonstrate

PMID: 19433090 [PubMed - indexed for M



Abstract

Similar articles

Metabolomic and genetic analyses of flavonol synthesis in Arabidopsis [Planta. 2009]

Characterization of flavonol synthase and leucoanthocyanidin dioxygenase [Plant Physiol. 1997]

The Arabidopsis TDS4 gene encodes leucoanthocyanidin dioxygenase [Plant J. 2003]

Review Multifunctional flavonoid dioxygenases: flavonol synthase [Phytochemistry. 2010]

Review New perspectives on proanthocyanidin biosynthesis [Phytochemistry. 2003]

See reviews...

Related information

Articles frequently viewed together

BioSystems

Gene

Gene (GeneRIF)

HomoloGene

Nucleotide (RefSeq)

Nucleotide (Weighted)

Protein (RefSeq)

Protein (Weighted)

PubChem Compound

PubChem Compound (MeSH Keyword)

PubChem Substance

PubChem Substance (MeSH Keyword)

Taxonomy via GenBank

UniGene

Protein

GEO Profiles

Cited in PMC

Look for the link to related genes and follow it.

ESTUDO DIRIGIDO

1. Introdução à Bioinformática
2. Bancos de dados em biologia: foco em alguns bancos de dados e ferramentas do NCBI

LEITURA RECOMENDADA

Griffiths AJ et al. Introdução à genética. Capítulo 14 Genomas e Genômica, pg: 447-480, 2016.



Illustration 3. An HP-9000 desktop workstation running the Unix-based system HP-UX. Image: Thomas Schanz//CC-BY-SA 3.0.