Noções de Bioinformática

Prof: Victor Martin Quintana Flores

Disciplina:
Fundamentos de Biotecnologia

O que é Bioinformática?

Bioinformática: Pesquisa, desenvolvimento ou aplicação de ferramentas e estratégias computacionais para expandir a utilização de dados biológicos, médicos, do comportamento ou da saúde, incluindo aquelas para obter, estocar, organizar, arquivar, analisar, ou visualizar tais dados.

Biologia Computacional: Desenvolvimento e aplicação de métodos teóricos e analíticos para análise de dados, modelagem matemática e técnicas de simulação computacional para o estudo de sistemas biológicos, comportamentais e sociais.

Segundo o NCBI

National Center for Biotechnology Information

Bioinformática é o campo da ciência no qual a biologia, a ciência da computação e a tecnologia da informação se misturam para formar uma única disciplina. O objetivo final deste campo e permitir a descoberta de novos conceitos biológicos bem como a criação de uma perspectiva global que permita o discernimento de princípios biológicos gerais.

Aplicações de Bioinformática

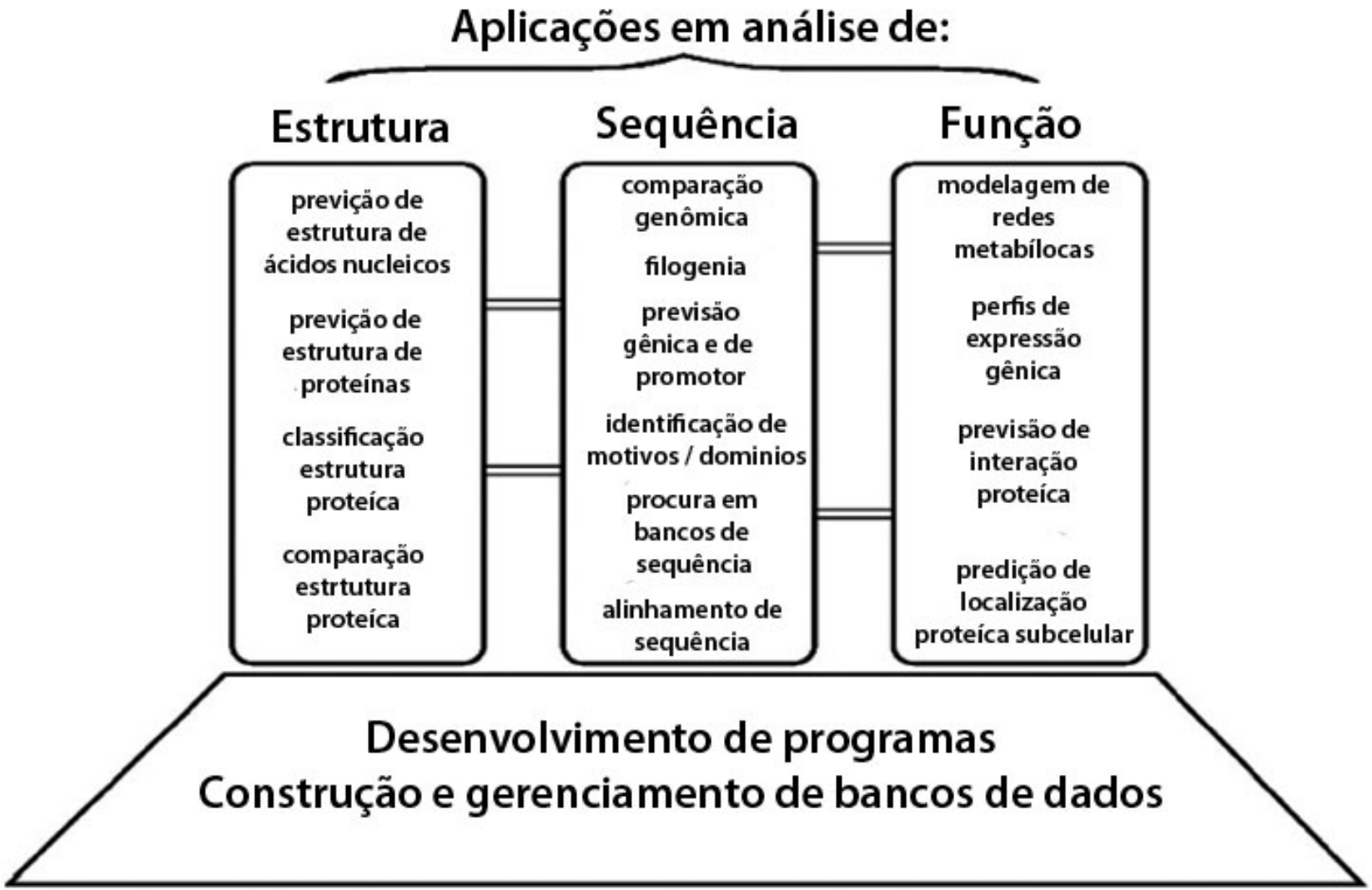
Há três importantes sub-disciplinas dentro da Bioinformática:

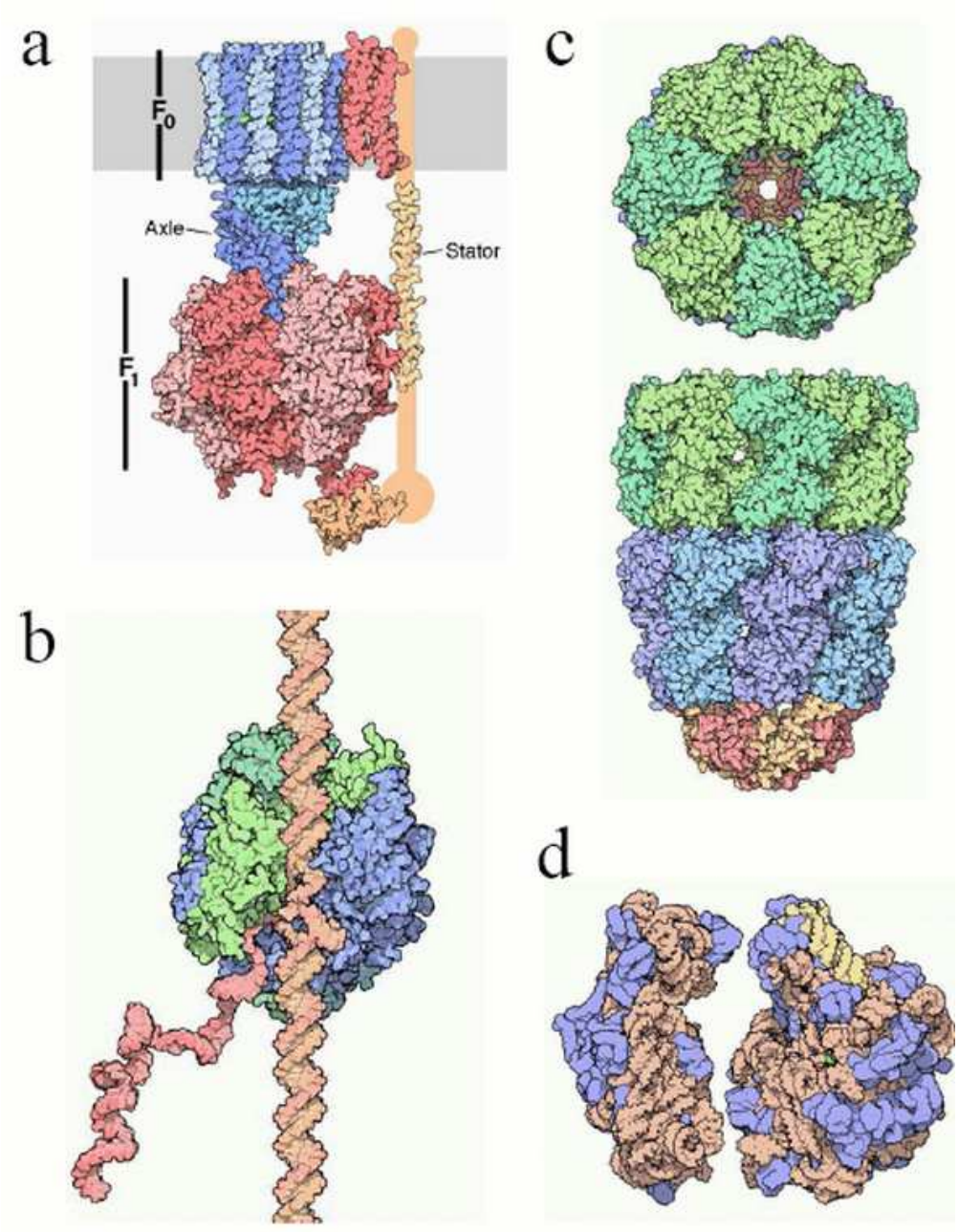
- o desenvolvimento de novos algoritmos e análises estatísticas com os quais inferir relações entre membros de grandes conjuntos de dados;

- a análise e interpretação de vários tipos de dados, incluindo seqüências de nucleotídeos e aminoácidos, domínios em proteínas, e estruturas protéicas;

- o desenvolvimento e implementação de ferramentas que permitam acessar e manipular de forma eficiente diferentes tipos de informação.

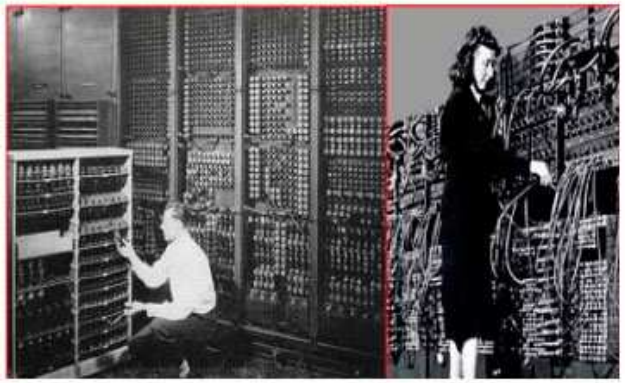
Aplicações de Bioinformática





Four molecular machines formed principally by proteins. Figures taken from the Molecule of the month section of the RSCB Protein Data Bank, we thank the RSCB PDB and David S. Goodsell, from the Scripps Research Institute, for kind permission to use them. **a)** *ATP synthase*: it acts as an energy generator when it is traversed by protons that make its two coupled engines rotate in reverse mode and the ATP molecule, the gas of the cell, is produced. **b)** *RNA polymerase*: it slides along a thread of DNA reading the base pairs and synthesizing a matching copy of RNA. **c)** *GroEL-GroES complex*: it helps unfolded proteins to fold by sheltering them from the overcrowded cellular cytoplasm.**d)** *Ribosome*: it polymerizes amino acids to form proteins following the instructions written in a thread of messenger RNA.

Breve histórico



ENIAC – 30 tn – 160 m²

Apple

Microsoft

National Center for Biotechnology Information (NCBI)

Estabelecer bancos de dados públicos

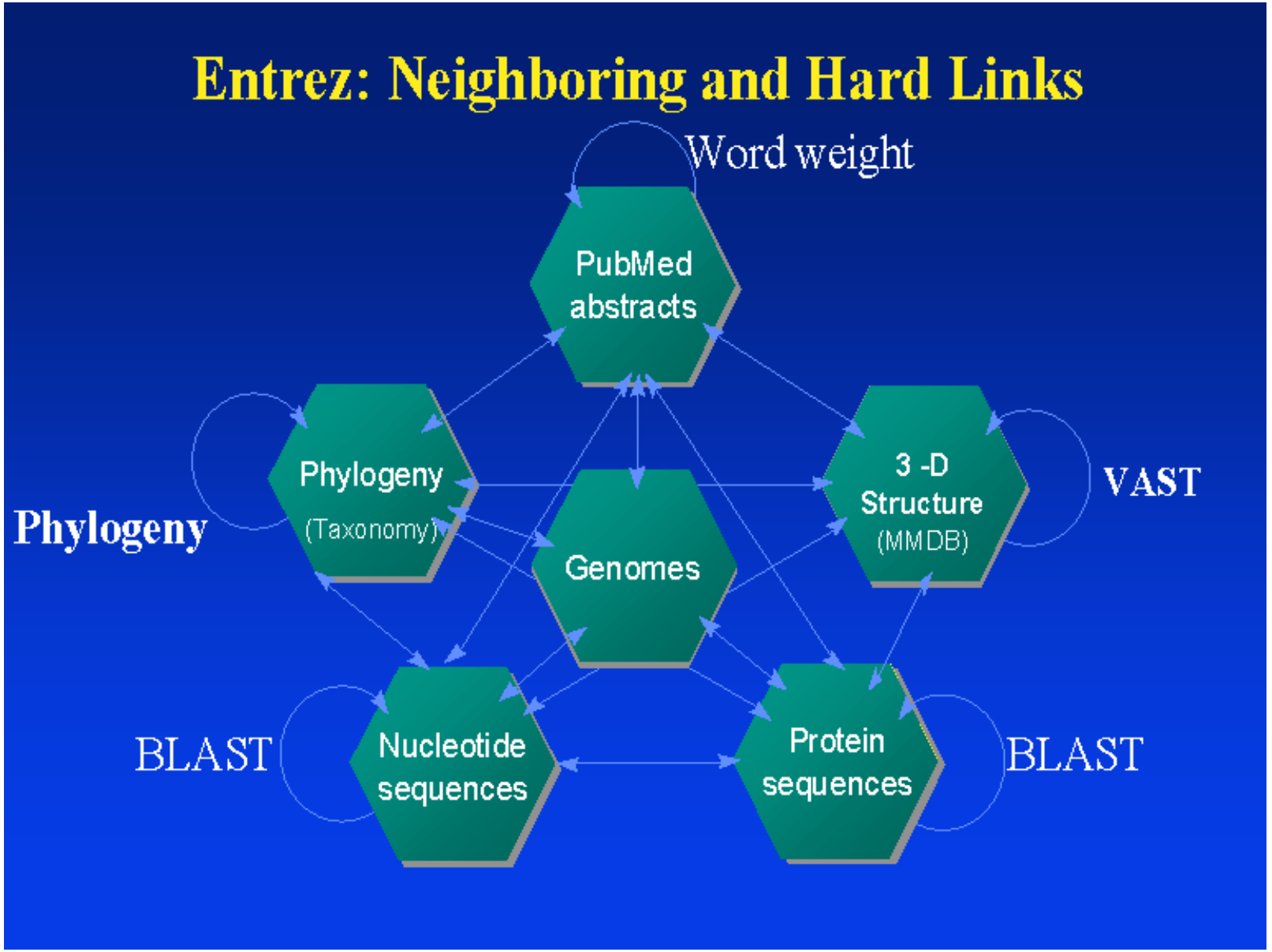
Desenvolver pesquisa em biologia computacional

Desenvolver ferramentas e software para análise de seqüências

Disseminar informações biomédicas

Bancos de dados e serviços disponíveis no NCBI

- GenBank largest sequence database
- Free public access to biomedical literature
  - PubMed free Medline
  - PubMed Central full text online access
- Entrez integrated molecular and literature databases
- BLAST highest volume sequence search service
- VAST structure similarity searches
- Software and Databases


Entrez: Neighboring and Hard Links
Word weight
PubMed abstracts
Phylogeny (Taxonomy)
Phylogeny
3 -D Structure (MMDB)
VAST
Genomes
BLAST
Nucleotide sequences
Protein sequences
BLAST

Tipos de Banco de Dados

- **Bancos de dados Primários**
  - Dados obtidos diretamente de seqüenciamento
  - Dados submetidos por pesquisadores
  - Conteúdo controlado pela pessoa que o submete
    - Exemplos: **GenBank, EMBL, DDJB, SNP, GEO**

- **Bancos de dados Derivados (ou Secundários)**
  - Construído a partir da base de dados primária
  - Padrões resultantes da análise dos primários
  - Conteúdo controlado por curadores (NCBI)
    - Exemplos: **Refseq, RefSNP, UniGene, NCBI Protein, Structure, Conserved Domain, SwissProt, Pfam**

Bancos de dados primários

GenBank

- Banco de dados de sequências de nucleotídeos gerados a partir do seqüenciamento de DNA
- Desde 1982 o número de bases dobra a cada 14 meses

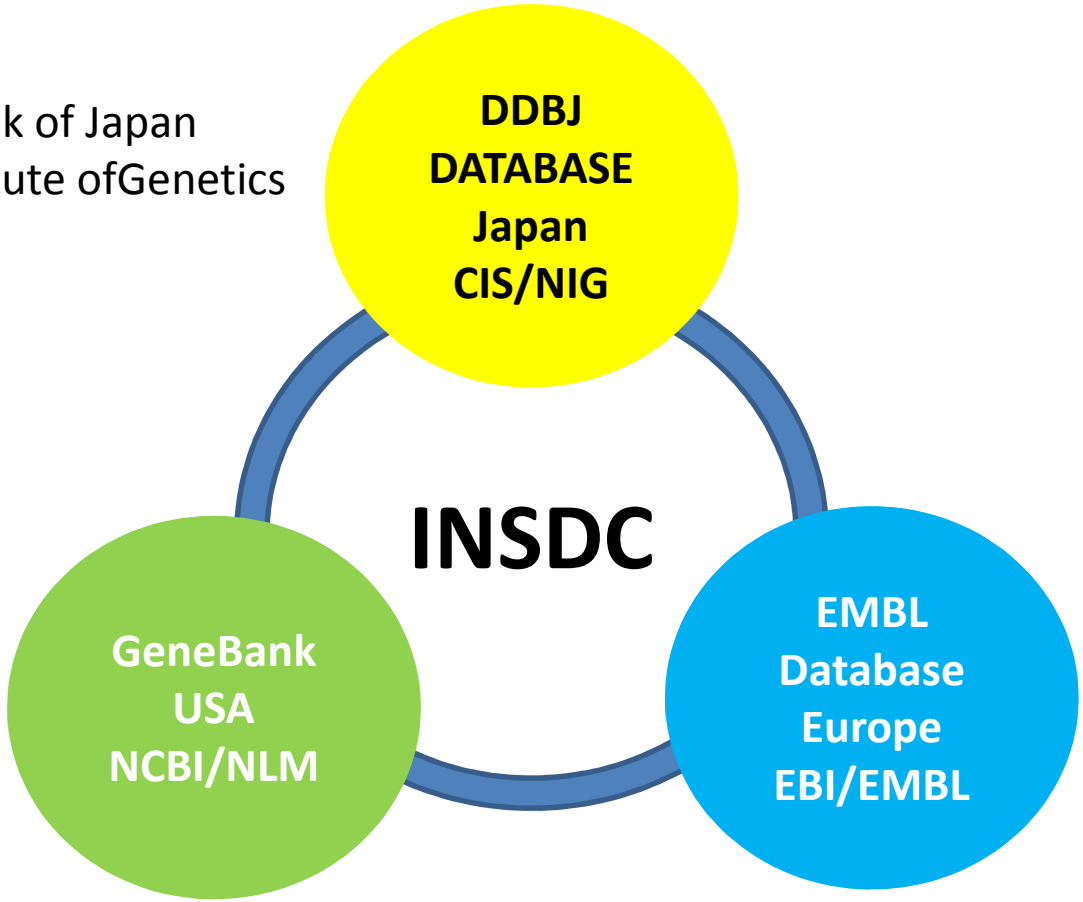
Modelo de Dados

- Bancos de dados de seqüências e as ferramentas de acesso do NCBI foram construídos a partir de um Modelo de Dados particular

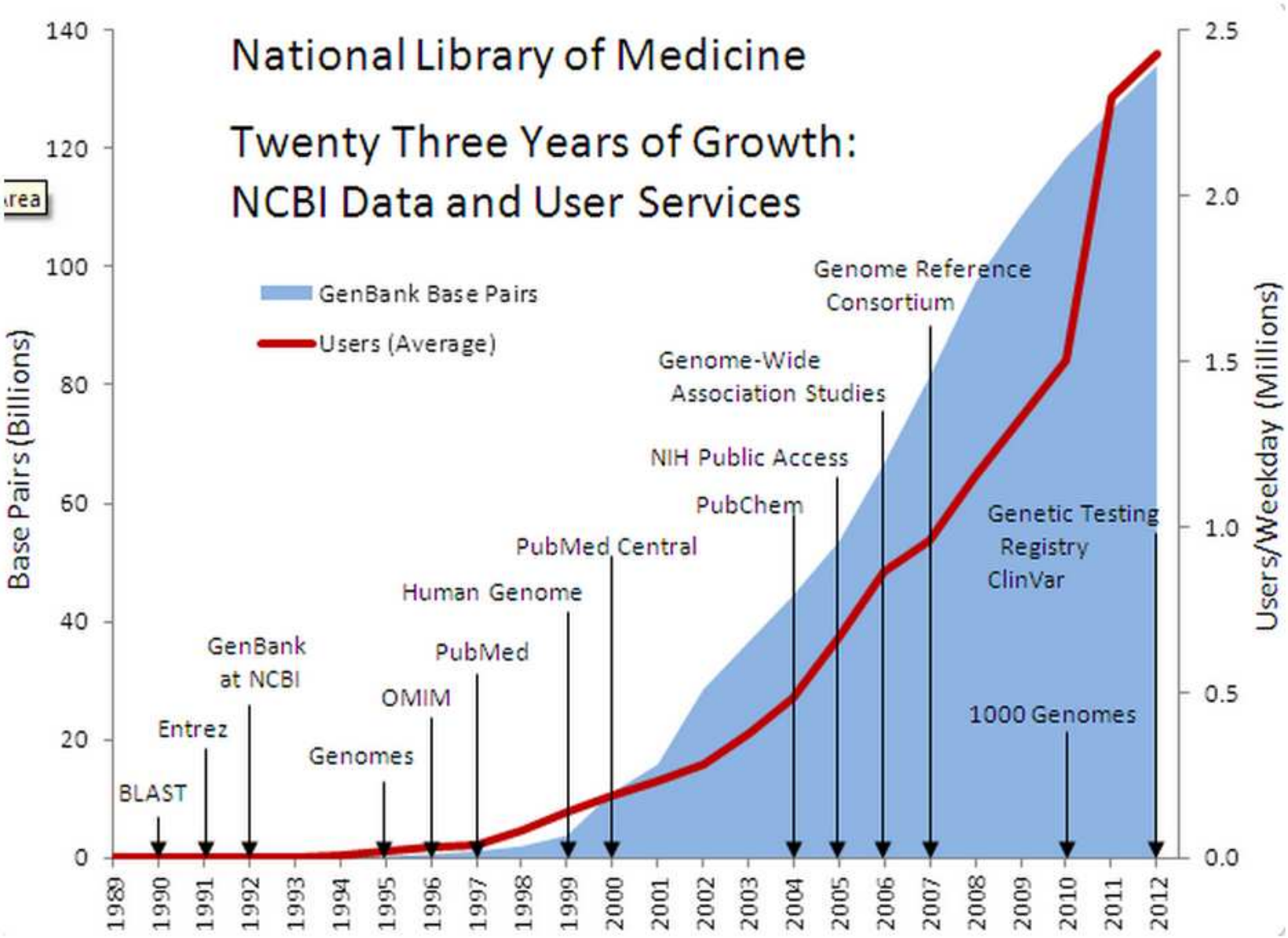
- Modelo de dados simples e poderoso o suficiente para agregar dados heterogêneos
  - sequências de nucleotídeos e aminoácidos
  - estruturas tridimensionais
  - publicações (Medline)

International Sequence Database Collaboration

DNA Data Bank of Japan
National Institute ofGenetics



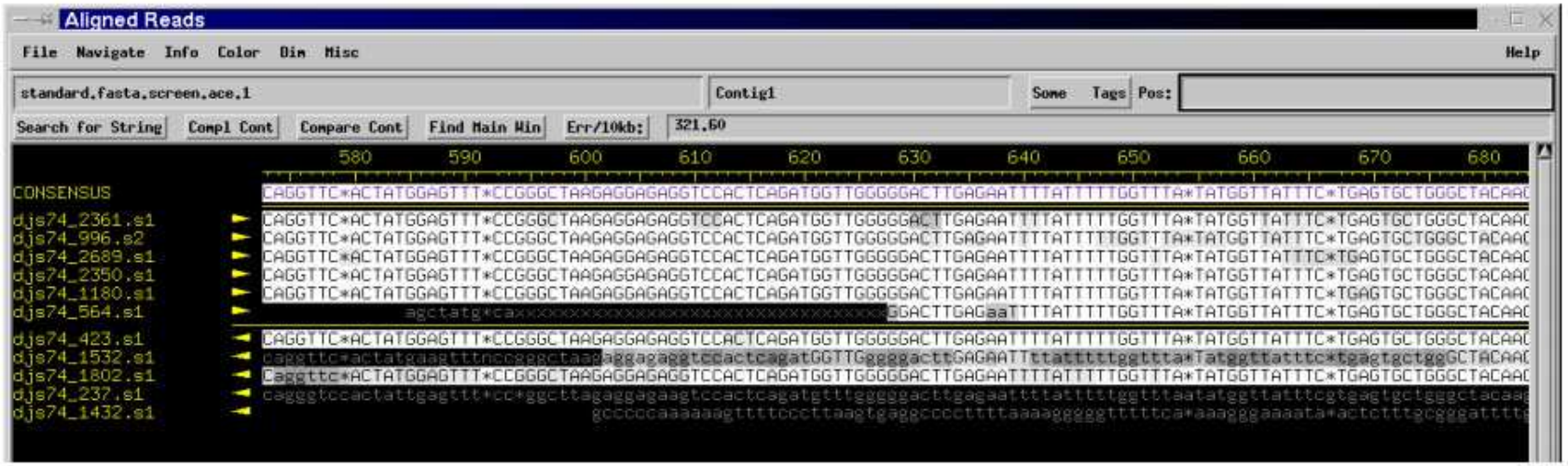
Crescimento das informações no GenBank



Exemplos Sequenciamento

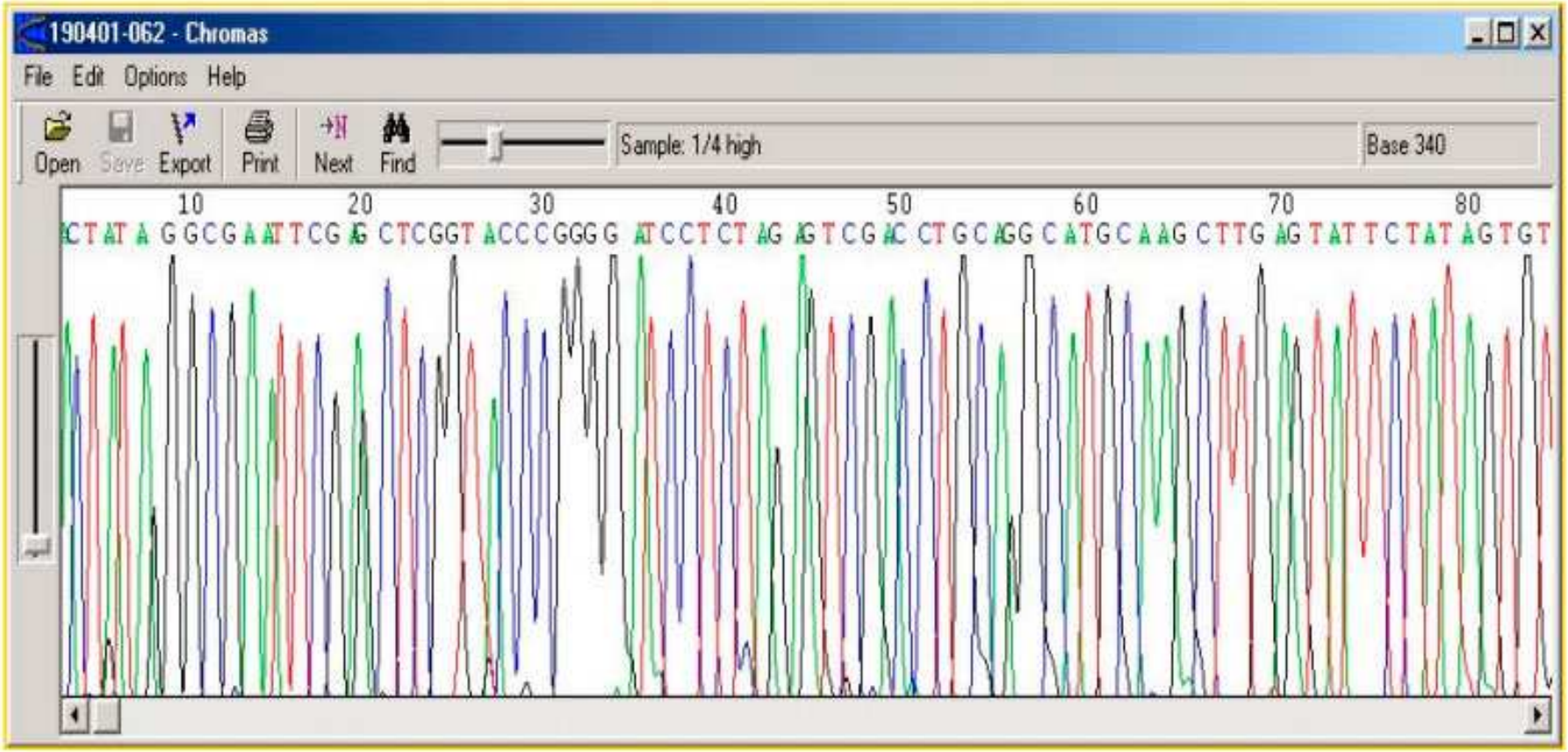
Phred/Phrap/Consed

- Phred – valores de qualidade para bases
- Cross-match – marcação
- Phrap e CAP3 – montagem
- Consed - visualização



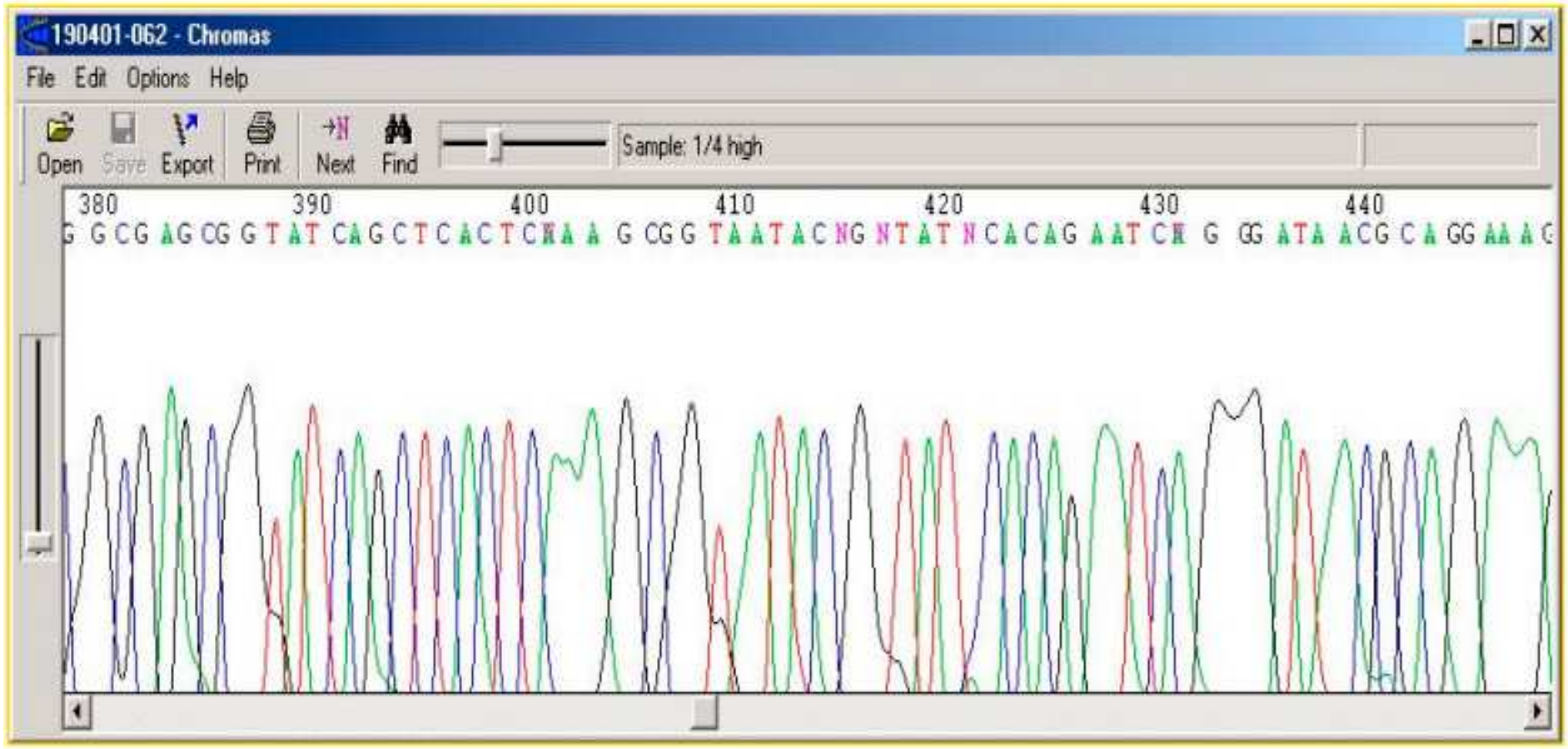
Exemplo eletroferograma

Região de baixa qualidade – baixa confiabilidade



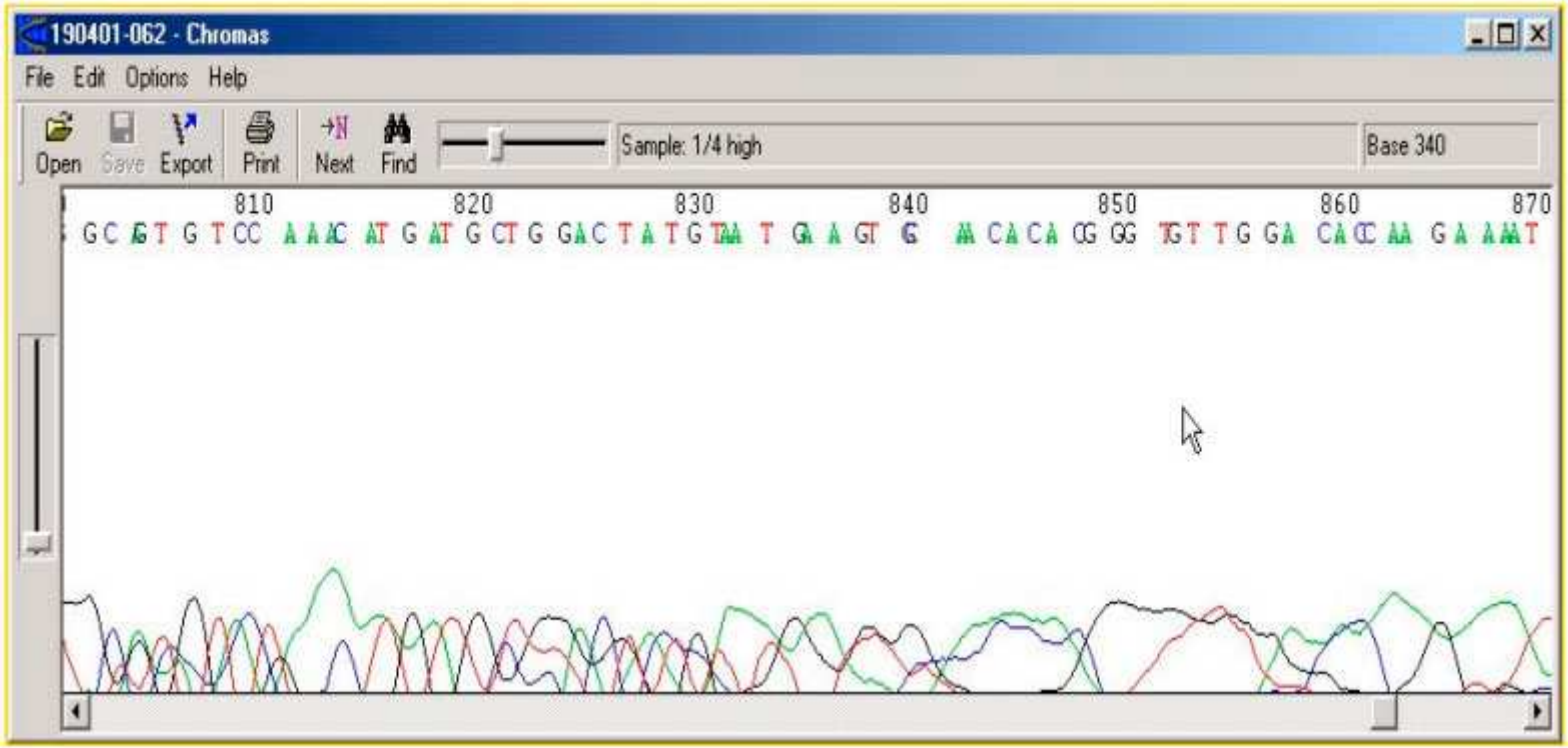
Exemplo eletroferograma

Região de qualidade média – poucas ambigüidades



Exemplo eletroferograma

Região de baixa qualidade – baixa confiabilidade



Bancos de dados e ferramentas do **NCBI**

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

BLAST - ferramenta de busca de homologia por alinhamento local

http://www.ncbi.nlm.nih.gov/BLAST

Phred, Phrap e Consed - ferramentas para análise da qualidade de seqüências e para montagem e visualização de contigs

http://www.phrap.org

COG - Cluster of Ortolog Groups - Bancos de dados filogeneticamente referenciado.

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

UniGene - Agrupamento de seqüências em consensos de genes.

http://www.ncbi.nlm.nih.gov/UniGene

LocusLink - ferramenta para recuperação de seqüências funcionais curadas.

http://www.ncbi.nlm.nih.gov/LocusLink

Gene Ontology Consortium - banco de dados genômicos para categorização dos genes de acordo com suas funções moleculares, processos biológicos e componentes celulares.

http://www.geneontology.org

Orchid BioSciences - empresa da área farmacogenômica

http://www.orchid.com

Celera - mega-empresa da área genômica

http://www.celera.com

ACT - Artemis Comparison Tool - comparação de genomas inteiros

http://www.sanger.ac.uk/Software/ACT

National Center for Genome Research (USA) - ferramentas de anotação

http://www.ncgr.org

Laboratório de Bioinformática da Unicamp

http://www.lbi.ic.unicamp.br

European Bioinformatics Institute - ferramentas e bancos de dados
http://www.ebi.ac.uk

The Biocomputing Service Group - várias ferramentas de análise genômica e anotação
http://genome.dkfz-heidelberg.de

TIGR - ferramentas para anotação e montagem final e visualização de genomas
http://www.tigr.org/software

GenScan - programa para predição de ORFs em um segmento genômico
http://genes.mit.edu/GENSCAN.html

ESTScan - programa para identificação de fase de leitura através do codon usage
http://www.ch.embnet.org/software/ESTScan.html

Núcleo de Bioinformática da UFMG - ferramentas simples de análise
http://www.icb.ufmg.br/~infobio