



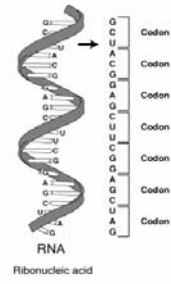
Bioinformatics at IU



```

. . . caaaaatagggttaatatgaatctcgatctccatttggttcatcgtattcaa
caacaagccaaaactcgtacaaatagaccgcaacttgcataaagaacaaggcttggg
cgagatatctctggaaaaactttcaagagcaactcaatcaacttctcgagcattgctt
gctcacaatattgagtcacaagataaaatcgccattttggccataataggaaacttgg
gttgttcatgaaacttgggtatcacaagatggtttaatgaccactgttcaagcaagact
acaaatggtgcaattggcaacttacaatctcgagcaatcaagtgccatttcaagcaaac
aatacagcccagcaagcagaatttatcctaaatcagccgatgtaaaaattctctcgtc
ggcgatcaagagcaatcagatcaaacattggaattgctcatcattgtccaaaattacaa
aaaattgtagcaatgaaatccaccattcaattacaacaagatcctcttcttgcactgg

```



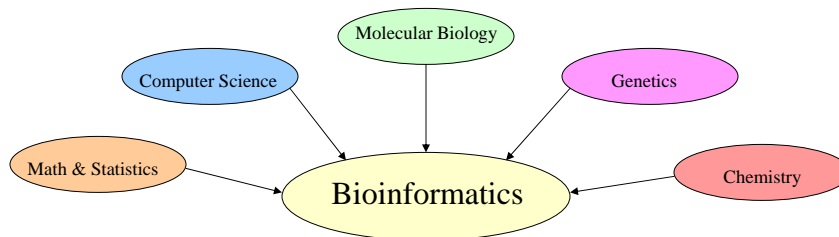
- Ketan Mane

Bioinformatics at IU

What is Bioinformatics?

Bioinformatics is the study of the inherent structure of biological information and biological systems

It is an amalgamation of systemic biological data (eg. genomes) with practical tools from computer science and other allied science streams.

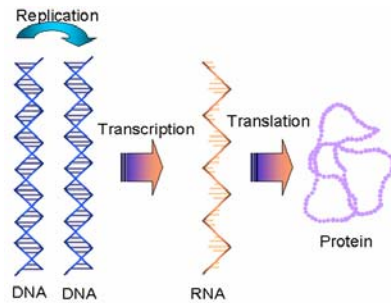


Bioinformatics at IU

Bioinformatics & Molecular Biology:

The aspect of Molecular Biology dealing with the DNA and its expression as proteins is dealt under the bioinformatics domain.

The experiments that monitor these transcription process generate huge amounts of data for further analysis.

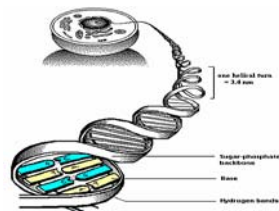


Bioinformatics at IU

Bioinformatics & Molecular Biology:

Bioinformatics helps in management (storage, retrieval and manipulation) of DNA, RNA and protein information from the perspective of molecular biological studies:

- Perform computing on information for DNA, RNA & protein sequence information (similarity searches, alignments)
- Study of evolutionary relationships between organisms through DNA and protein sequence homologies (similarities)
- Based on earlier repository of data, obtain information on structure-function, evolution of genes, and organisms
- Design of new experiments involving gene expression studies



Bioinformatics at IU

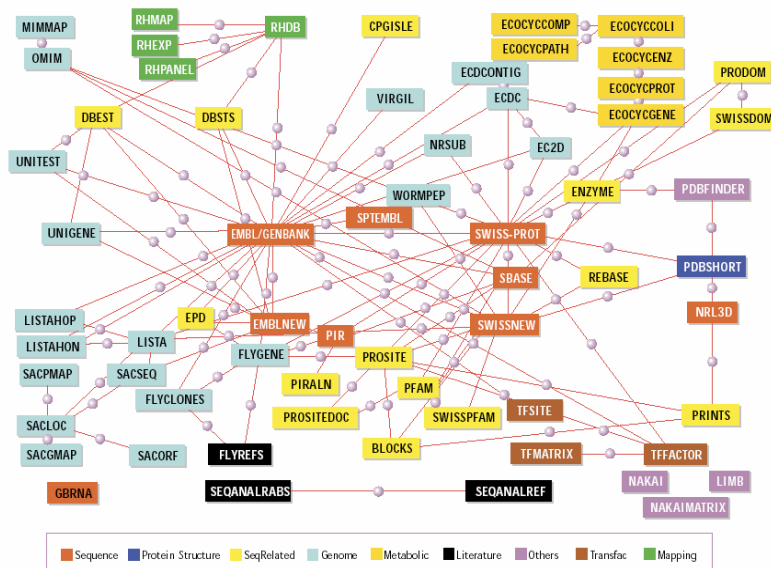
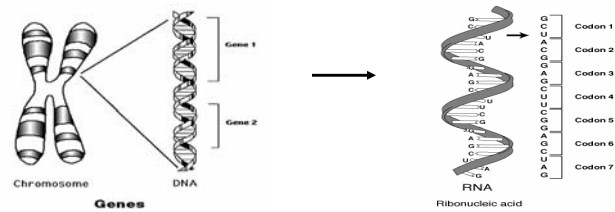
Bioinformatics and Genomics:

Genes influence our physiological and pathological processes. Individual variation in the gene sequence is responsible for diseases. Knowledge of these sequences would contribute towards possible solutions to control these diseases.

Genomics deals with the sequencing of the genetic information, deciphering the exact sequence of lettered bases which compose a gene

Experiments are performed to study this gene sequences and how they influence the disease. Recent advances in instrumentation has led to explosion of sequence information.

Human Genome Project is working towards obtaining sequence information of all the genes in human cells and creating the complete map of genes. This would help to uncover the genetic mysteries at the heart of diseases like cancer. This would lead to better understanding of creating new diagnostic tools and improved treatments.



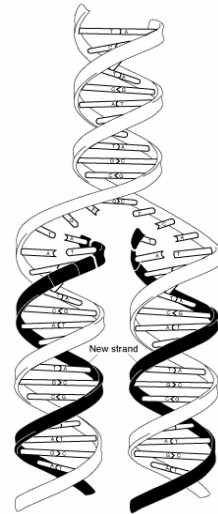
Source:- Don Gilbert

Bioinformatics at IU

Bioinformatics and Genomics:

Bioinformatics assists genomic studies through:

- Sequence storage and retrieval of gene sequence information
- Simple pattern searching in sequences
- Comparing sequences through dot matrix method for predicting the functions of the gene.
- Searching sequence databases for similar sequences



Bioinformatics at IU

Bioinformatics and Chemistry:

The branch of chemistry that is related to bioinformatics is “Proteomics” which deals with the study of proteins in the cell.

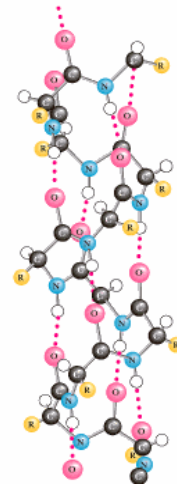
Primary structure of proteins is the linear sequence of amino acids that are bound together by peptide bonds. Change in a single amino acid in a critical area of the protein can alter biologic function as is the case in sickle cell disease and many inherited metabolic disorders.

“Proteome” → Quantitative protein expression pattern of a genome under defined conditions

Proteomics involves the identification of proteins in the body and the determination of their role in physiological and patho-physiological functions.

Proteomics research offers promises for:

- novel drug discovery via the analysis of clinically relevant molecular events
- Process optimization for recombinant proteins
- Diagnostics (e.g. tumor markers)



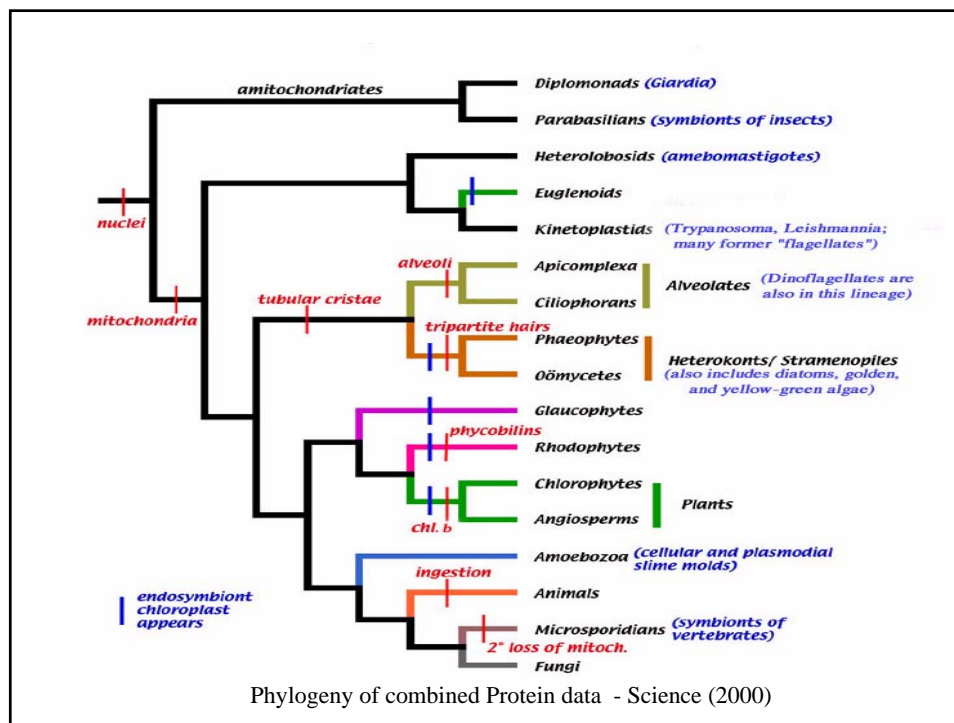
Bioinformatics at IU

Bioinformatics and Math & Statistics:

Math & Statistics are make its contribution to this field by providing methods for development of the bioinformatics tools.

A background knowledge in this area is required :

- To design experiments, which aid the understanding of gene function, genetic networks, genetic regulation and control pathways, protein structure and genome organization
- To develop and evaluate molecular models
- In maintaining correlation between new and existing data
- To develop data mining techniques for discovering patterns and relationships in large complex data sets
- To determine extent of variation in protein structures belonging to the same phylogeny

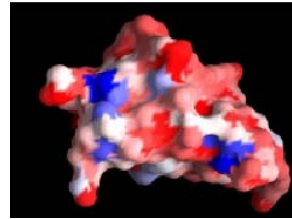


Bioinformatics at IU

Bioinformatics and Computer Science:

Computers are powerful devices for understanding any systems that can be described in mathematical way. This field contributes to the bioinformatics area by:

- Developing algorithms to determine gene sequence from raw experimental data
- Using image processing techniques to expedite the process of mapping the genomic data
- Exploring the potentials of parallel computing for data collection and analysis
- Providing tools for developing communication and information transfer over the internet
- Introducing intelligent designs for data format and databases



Bioinformatics at IU

Bioinformatics and Computer Science:

- Developing user-interfaces that bring together different tools to query datasets
- Assist in developing interactive applications to give a good insight on data



Bioinformatics at IU

Research in Molecular Biology:

Indiana Molecular Biology Institute at IUB is involved with research in nucleic acid technologies, regulation of gene expression, cell, developmental, genetic and molecular biology, molecular evolution, microbiology, and biochemistry

Some of research involvement in this areas include:

- *Genetics And Molecular Biology Research*
Study of nuclear receptors in their behavior towards different hormones. Effects of hormones are being studied through in *Drosophila*, whose genome sequence is available
- *Molecular Evolution Research*
Manipulating genetically, the molecular development of egg in *Drosophila*, to address issues concerning the organization and evolution of pattern via signal transduction
- *Gene Expression Study Research*
Mechanistic studies of DNA mismatch and lesion processing, An attempt to solve the enigma about how some proteins recognize mismatches and corrects them. Also it recognizes certain types of DNA damage and attempts to process them



Bioinformatics at IU

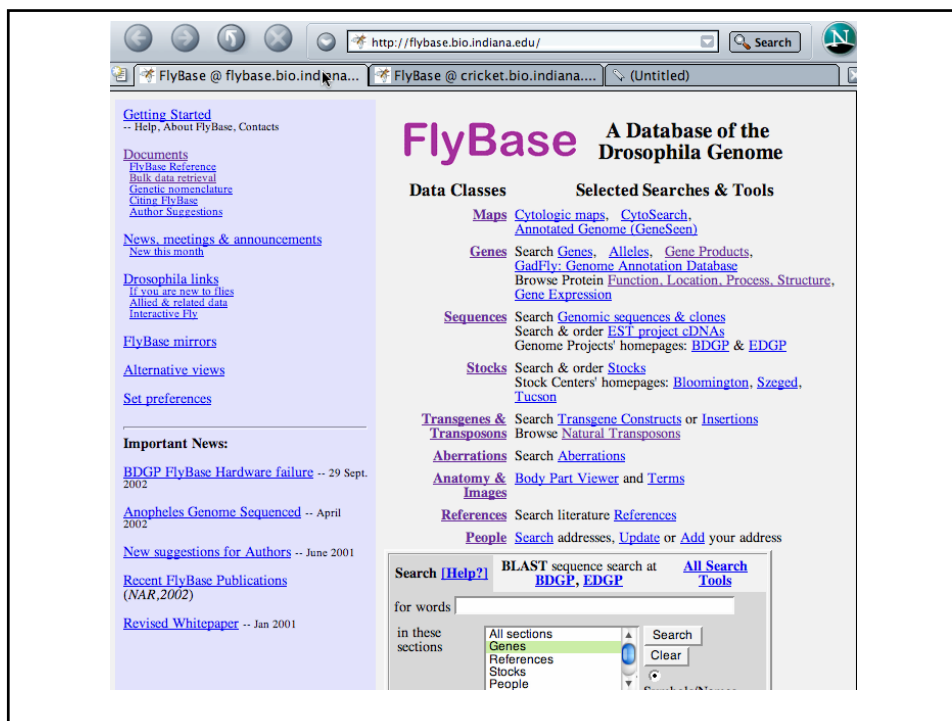
Research in Genomics:

The CGB Genomics laboratory at IUB is equipped for high throughput sample handling, microarray fabrication & analysis, and single nucleotide polymorphism (SNP) determination.

The Genomics Group at IUB is involved in a number of projects both independently and in collaboration with IUB faculty:

- *Sex Genes in Drosophila*
An ongoing project in the genomics focuses on sex determination in *Drosophila*
- *Sunflower Genomics*
Deals with identifying single nucleotide polymorphisms (SNPs) in the sunflower genome
- *FlyBase Project*
Maintains *Drosophila* Genome sequence information and also serves as *Drosophila* Genome Information Center
<http://fly.bio.indiana.edu/>





Bioinformatics at IU

Genome Data Objects

Contents	Entries
Literature References	140,000
Gene variants	112,000
Genome features	50,000
Genes	40,000
Transgene constructs	37,000
Chromosome aberrations	16,000
Fly Stocks	15,000
Drosophila Researchers	6,600
Total	416,600

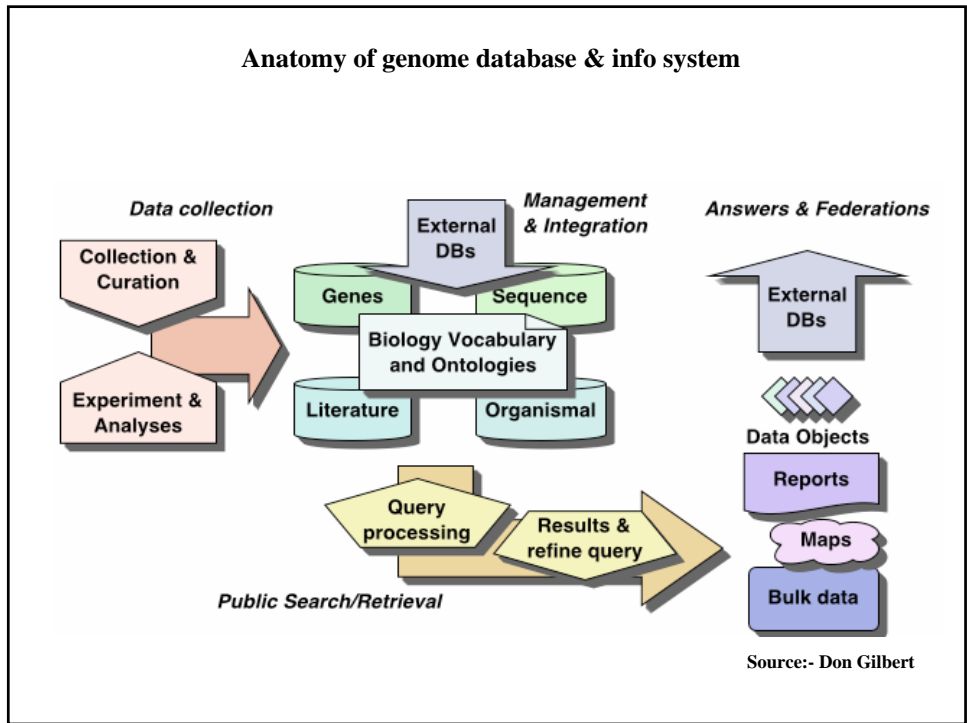
Drosophila genome, FlyBase, Sept. 2002

Source:- Don Gilbert

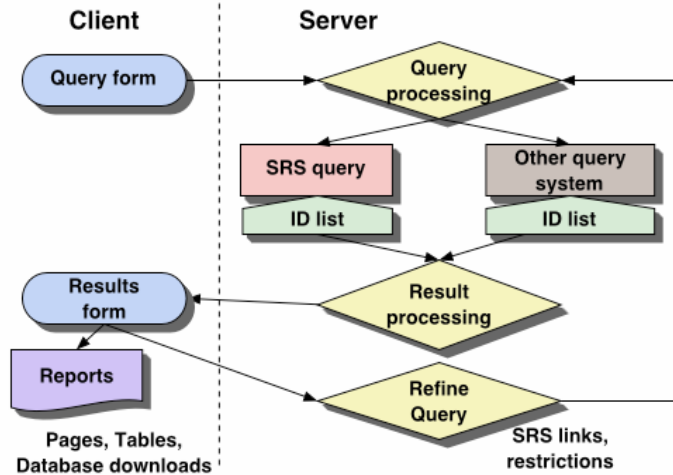
The screenshot shows the euGenes website interface. On the left, there is a table listing various organisms and their available genes:

Organism	Available genes
Fruitfly Genes (Drosophila, melanogaster)	25,728
Human Genes (Homo sapiens)	53,210
Mouse Genes (Mus musculus)	36,433
Mosquito Genes (Anopheles gambiae)	12,687
Weed Genes (Arabidopsis thaliana)	26,819
Worm Genes (Caenorhabditis elegans)	22,705
Yeast Genes (Saccharomyces cerevisiae)	7,222
Zebrafish Genes (Danio rerio)	1,583
All Genes summaries	

The main content area is titled "Human Genes: Genomic Information for Homo sapiens". It includes a search bar with a dropdown menu set to "Everything", a "Search Help?" link, and a checkbox for "wildcard * to words". Below the search bar, there are links for "Internet services for Human Genes" and "Human Genes Data files".



FlyBase/euGenes Query System



Source:- Don Gilbert

The screenshot shows a web browser window displaying the IUBio Archive website. The address bar shows <http://iubio.bio.indiana.edu>. The page content includes:

- IUBio Contents** section with links to:
 - euGenes**: Genome information system for eukaryotic organisms, including genome maps, gene homologies, product and related information.
 - Software**: Software for molecular biology and related areas, organized by categories (alignment, evolution, etc.) and by computer system (macintosh, ms dos/win, unix, java). See also this [descriptive list](#), or [ePrints database](#) of molecular biology software.
 - DNA & protein sequences**: Search and fetch DNA and protein sequences, including > [SRS](#), Sequence Retrieval System > Fetch bulk data from [Bio-Mirror of large data sets](#) including GenBank, EMBL and DDBJ and other biosequence data, updated daily.
 - Bionet News**: Bionet news groups archive, including molecular methods, software, and others, with a search of current or previous years back to 1990.
- Quick links** section:
 - [SeqPup](#) - biosequence editor
 - [Readseq](#) - biosequence conversions ([web form](#))
 - [SeWeR](#) - Sequence Analysis using Web Resources
 - [FlyBase](#) - Drosophila Genome Information service
 - [Phylodendron](#) - phylogenetic tree drawing
 - IUBio Archive [by FTP](#)
- IUBio-Archive** logo and description:

This is an archive of biology data and software, established in 1989 to promote public access to freely available information, primarily in the field of molecular biology.

Here you find ways to browse, search and fetch molecular data, software, biology news and documents, as well as links to remote information sources in biology and elsewhere. To learn more, read [about IUBio Archive](#).
- News** section:
 - [BioGrid and Bio-Data warehouse efforts](#), 2002
 - [Bio-Software ePrints database](#), 2002
 - [Internet robots blocked](#) from search of Bionet news archive, 2002
- Search** section:

Search for

in [Software abstracts](#)
[DNA sequences](#)
[Protein sequences](#)
[Bionet News](#)
- Footer: [Genome Informatics jobs at IU Biology](#), Nov. 2002

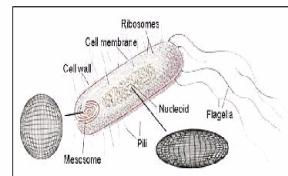
Bioinformatics at IU

Research in Chemistry at IU:

Research opportunities span nearly all aspects of structure, function, and dynamics of proteins and other biomolecules.

Some of the active ongoing research are as follows:

- Structural Studies of Protein-Protein Interactions in GroEL-Mediated Protein Folding Process and Host-Pathogen Communications
- Understanding the role of RNA: protein interactions in development, specifically on how proteins modulate the structure of RNA to control translation
- Developing comprehensive model of the living cell that accounts for genomics, proteomics and metabolics. This computational model would be used to accelerate drug discovery, optimize treatment strategies, and understand the fundamental workings of the living cell



Bioinformatics at IU

Research involving Computer Science:

- *BAG: A Graph Theoretic Sequence Clustering Algorithm*
BAG is a graph theoretic sequence clustering algorithm implemented in C++ with the LEDA package. This is used to compare many different set of genome
- *PlatCom: A Platform for Computational Comparative Genomics*
PlatCom is a system for the comparative analysis of multiple organisms. It involves 3 components: databases of biological entities; databases of relationships among entities; and a suite of mining applications over the databases.
- *Fast String Pattern Matching Algorithms*
Developing of new string matching algorithms.

```
. . . caaaaaataggggtaaatatgaatctogactccatttggctcatgtattcaa  
caacaagccaaaactcgtacaaatagaccgcaacttcgctataaagaacacggcttgg  
cgagatctcttggaaaaacttcaagagcaactcaatcaacttctogagcattgett  
gctcacaatttgacgtacaagataaaatcgccatttggccataataggaaagttgg  
gttggctatgaaacttctggatcaaaagatggtttaatgaccaactgttcacgcaacgact  
acaatcgttgacattgacaccttacaattogagcaatcacagtgcctatttacgcaacc  
aatcacgcccagcaagcagaatttatcctaaatcacgccgatgtaaaaattctctcgtc  
ggcgatcaagagcaatacgtacaaacattggaattgctcatcattgtccaaaattacaa  
aaaattgtagcaatgaaatccaccattcaattacaacaagatcctcttcttgcacttgg
```

Thank You !!!