Bioinformatics and Computational Biology

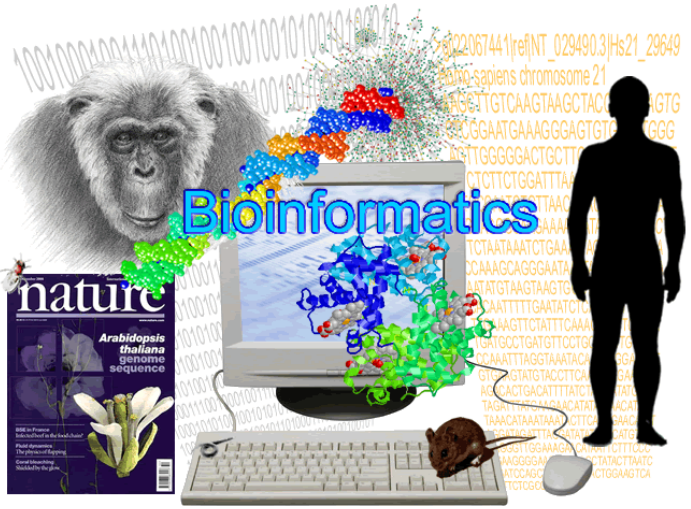
(Opportunities & Challenges to solve complex problems)

Rakesh Kaundal

Bioinformatics & Computational Biology Laboratory

(http://bic.okstate.edu/)

Department of Biochemistry & Molecular Biology



Bioinformatics

Why it's useful…

- All of the information needed to build an organism is contained in its DNA. If we could understand it, we would **know how life works**.
  - Preventing and curing diseases like cancer (which is caused by mutations in DNA) and inherited diseases.
  - **Curing infectious diseases** (everything from AIDS and malaria to the common cold). If we understand how a microorganism works, we can figure out how to block it.
  - Understanding genetic and **evolutionary relationships** between species
  - Understanding **genetic relationships** between humans. Projects exist to understand human genetic diversity
  - Similarly, other Eukaryotes are being sequenced including plants, e.g. to understand **plant diseases**, their **tolerance under stress conditions** *etc*.
  - Prokaryotes, **Metagenome** sequencing…….

Resistant **Susceptible**

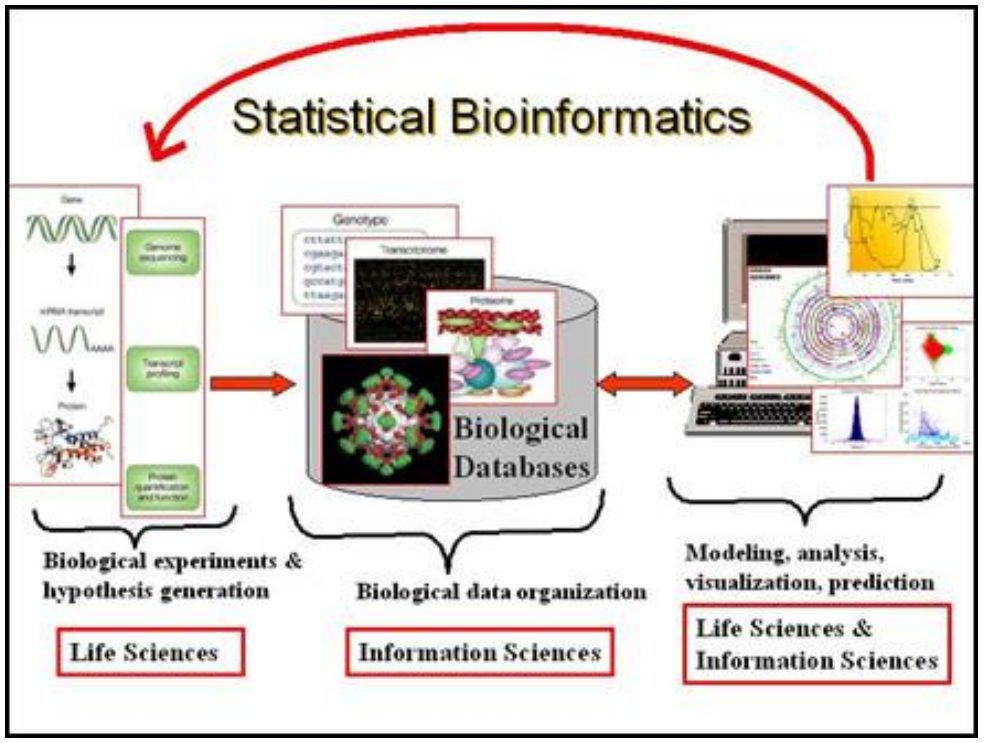
Abiotic stress

Large-scale sequence analysis

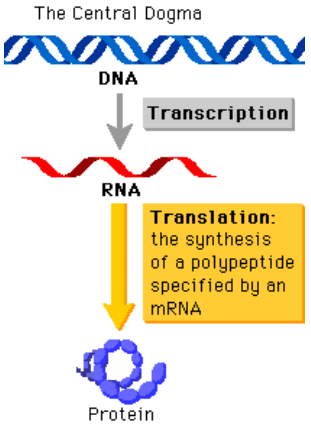
(Processing and Assembly)

Data Processing (e.g. EST to Unigene):

- EST cleaning (vector, adapter seqs etc.)
- Assembly (Tentative Consensus, Singletons)
- Unigenes
- Annotation



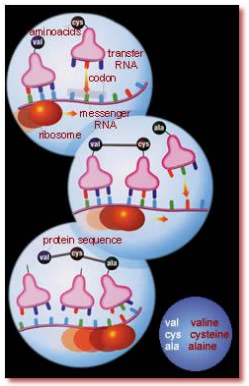
Large-scale sequence analysis



(Annotation: adding biological information to sequence)

- **Annotation by homology (BLAST)**
  - requires a large, well annotated database of protein / nucleotide sequences

- **Annotation by sequence composition**
  - simple statistical/mathematical methods



- **Annotation by sequence features, profiles or motifs (InterProScan, PFAM)**
  - requires sophisticated sequence analysis tools

- **Annotation by Artificial Intelligence**
  - Identify ‘known’ from the ‘unknowns’ based on training the features from known samples
  - computational tools for prediction

What if there is No homology in the reference database or NO HITS?

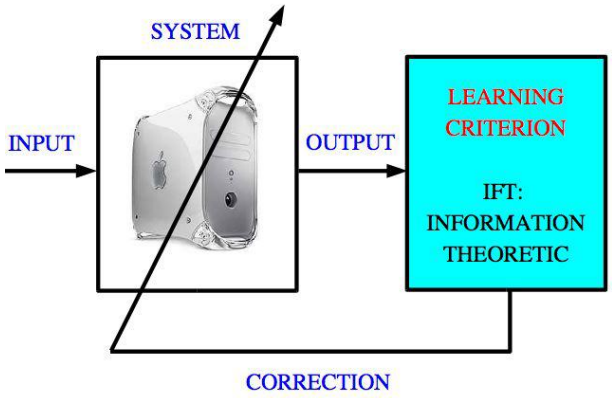
(***most of the time, this is the case……***)

Computational Intelligence

(offshoot of Artificial Intelligence)

- Machine Learning: an iterative *process* where a computer can learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks T, as measured by P, improves with experience.

  - Artificial Neural Networks (ANN)
    - Self-Organized Maps (SOMs)
  - Bayesian statistics
  - Decision tree learning
  - Support Vector Machines (SVM)

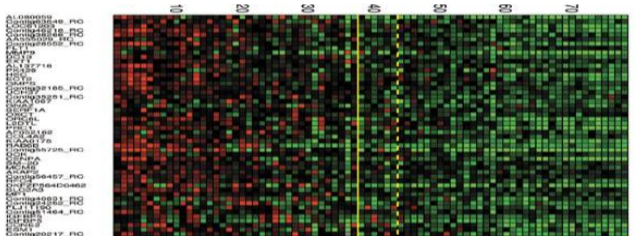


Modeling of Biological Systems

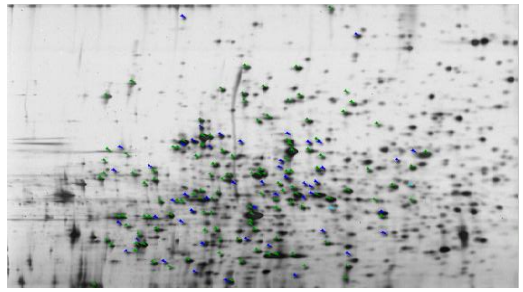
‘Omics’ technologies

DNA ⟷ **Genomics** ⟷

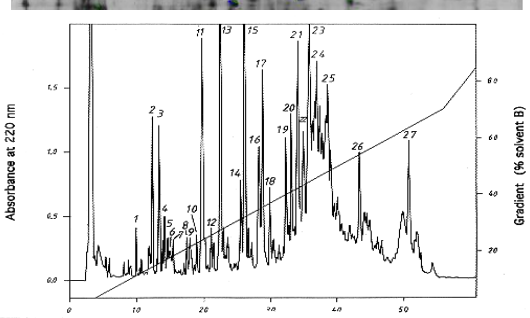
```
Q        366 TAACCTTGAATAATCACCACAGCAAAAATATTTTACCATTAAAAAATCAGTCTTAGAAAT
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
T   29069186 TAACCTTGAATAATCACCACAGCAAAAATATTTTACCATTAAAAAATCAGTCTTAGAAAT

Q        426 GATTTATCTATTTTAAGATTTTAATCTAATTTAGTGAATCAAAATTTTTAATCAGTGTTG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
T   29069246 GATTTATCTATTTTAAGATTTTAATCTAATTTAGTGAATCAAAATTTTTAATCAGTGTTG

Q        486 ATGGCAAATCTGCAAAAGGTATTTGTACTTTTACATAAGGATCTTAGAAATATGGTATAA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
T   29069306 ATGGCAAATCTGCAAAAGGTATTTGTACTTTTACATAAGGATCTTAGAAATATGGTATAA

Q        546 AAACAATTTTGTGCTAGATGACAAGCAATTTTGAGCTTTTTCTA
             ||||||||||||||||||||||||||||||||||||||||||||
T   29069366 AAACAATTTTGTGCTAGATGACAAGCAATTTTGAGCTTTTTCTA
```



mRNA ⟷ **Transcriptomics** ⟷

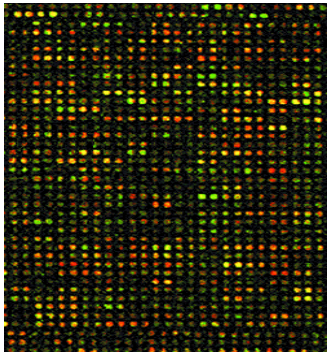


Protein ⟷ **Proteomics** ⟷



Metabolite ⟷ **Metabolomics** ⟷

Artificial Intelligence in Bioinformatics



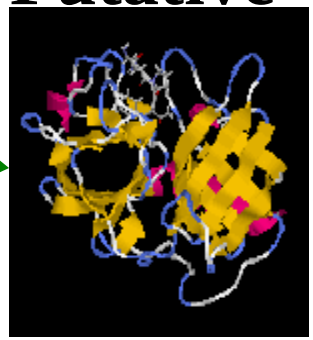
mRNA
expression data

Subcellular
Localization
Prediction

Metabolic Networks
(identify pathways)

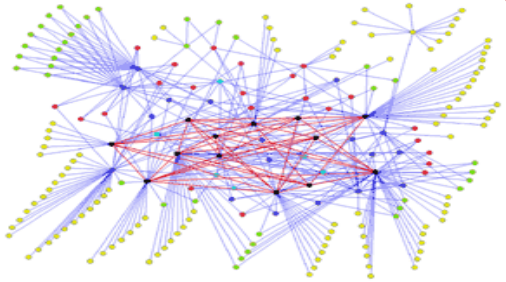
Computational
Epidemiology
(Disease prediction)

Putative



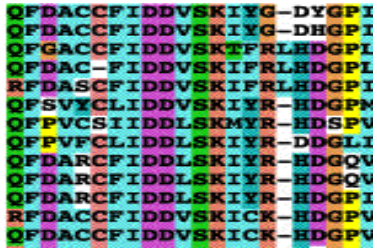
Function

Host-microbe
interactions



Gene Regulatory
Networks

Sequence signatures
(gene, protein)



Thank you for your kind attention!

http://bic.okstate.edu/ or http://biochemistry.okstate.edu/