

Tools and Algorithms in Bioinformatics

GCBA815, Fall 2015

Week1: Introduction to Bioinformatics

Babu Guda, Ph.D.

Department of Genetics, Cell Biology & Anatomy

Bioinformatics and Systems Biology Core

University of Nebraska Medical Center

Fall, 2015

GCBA 815

Topics:

- Class introduction
- Course details – syllabus, tests and grading
- Introduction to Bioinformatics

Course Page:

<http://www.unmc.edu/bsbc/education/courses/gcba815/>

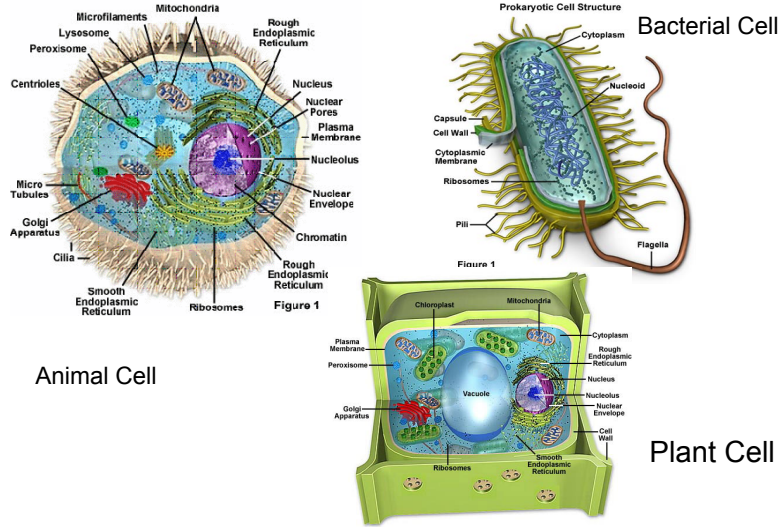
Online resources:

<http://www.unmc.edu/bsbc/education/courses/gcba815/resources.html>

Fall, 2015

GCBA 815

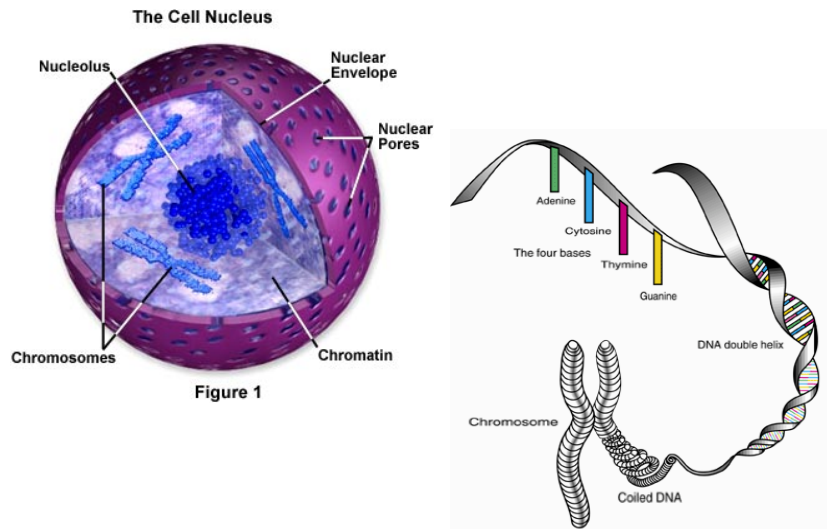
Eukaryotic and Prokaryotic Cells



Fall, 2015

GCBA 815

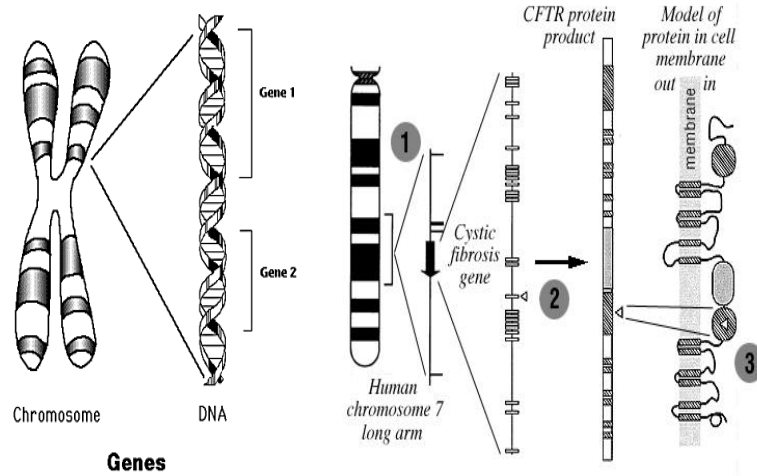
Nucleus-Chromosome-DNA



Fall, 2015

GCBA 815

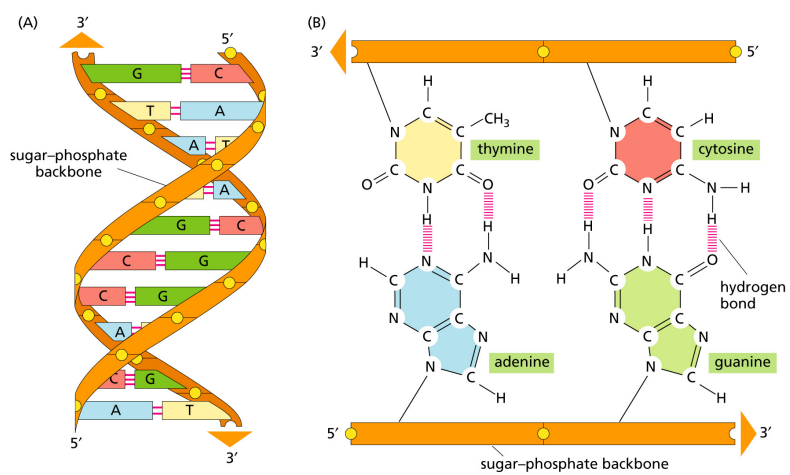
Gene mapping



Fall, 2015

GCBA 815

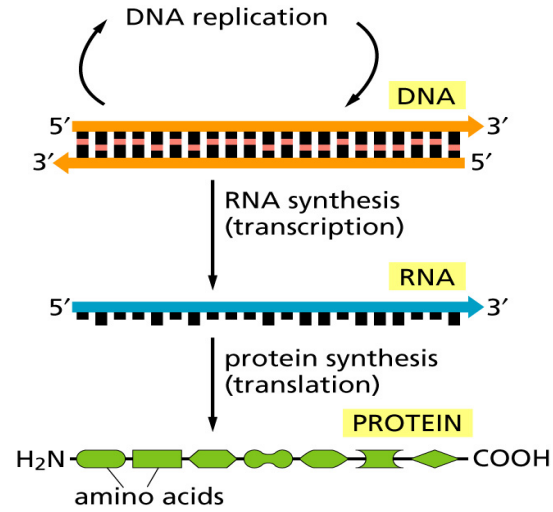
Nucleic Acids and Nucleotides



Fall, 2015

GCBA 815

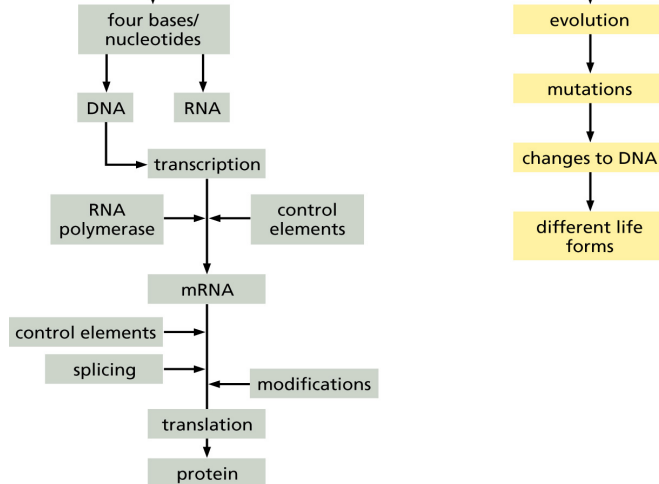
The Central Dogma of Molecular Biology



Fall, 2015

GCBA 815

THE NUCLEIC ACID WORLD

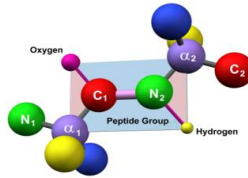
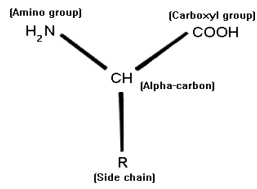


Fall, 2015

GCBA 815

Amino Acids The Building Blocks of Proteins

- **Twenty types of amino acids.**
- **All residues have main polar chain atoms (N and carbonyl). The central carbon atom is called the C α -atom and is a chiral center.**



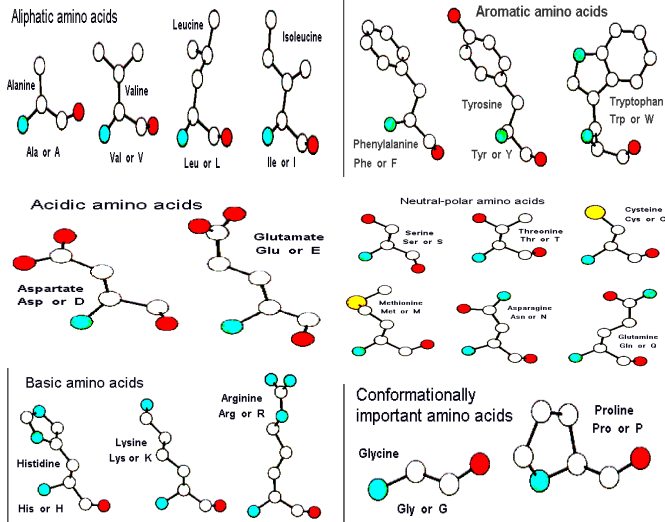
Alanine Ala A 90
 Cysteine Cys C 105
 Aspartic Acid Asp D 117
 Glutamic Acid Glu E 140
 Phenylalanine Phe F 193
Glycine Gly G 64
 Histidine His H 159
 Isoleucine Ile I 165
 Lysine Lys K 170
 Leucine Leu L 165
 Methionine Met M 168
 Asparagine Asn N 128
 Proline Pro P 124
 Glutamine Gln Q 150
 Arginine Arg R 194
 Serine Ser S 94
 Threonine Thr T 120
 Valine Val V 139
Tryptophan Trp W 233
 Tyrosine Tyr Y 198

- **Groups of Amino Acids based on their physico-chemical properties**
- **Tiny, Small, Aliphatic, Aromatic, Hydrophobic, Acidic, Basic, Polar, Charged**

Fall, 2015

GCBA 815

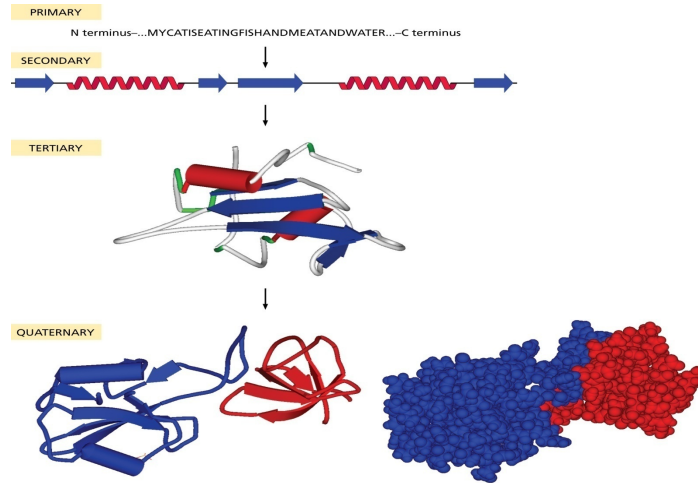
The 20 natural amino acids



Fall, 2015

GCBA 815

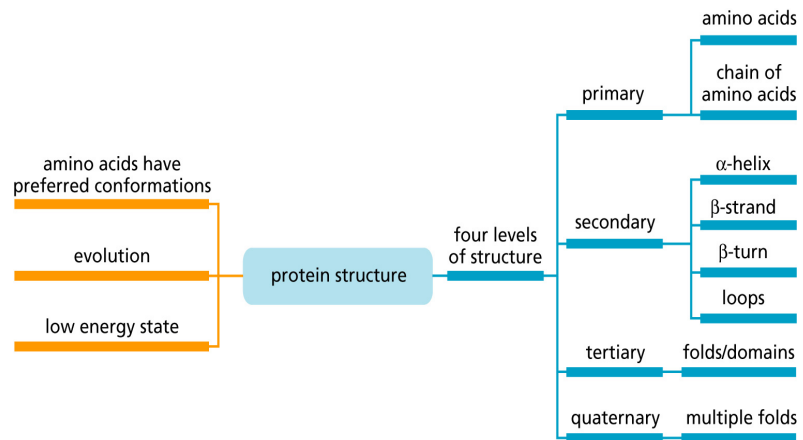
Hierarchical Levels of Protein Structure



Fall, 2015

GCBA 815

Summary of Protein Structure

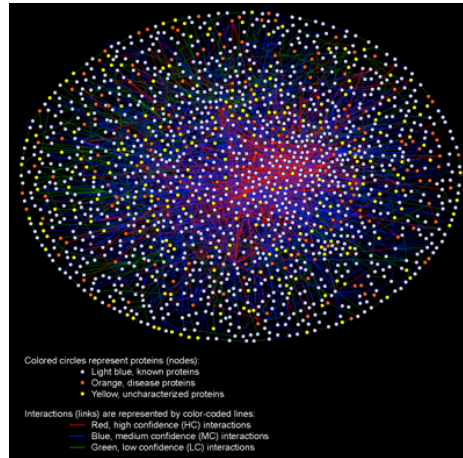


Fall, 2015

GCBA 815

Human Protein Interaction Network

(Ulrich et al., 2005, Cell 122:957-968)



Fall, 2015

GCBA 815

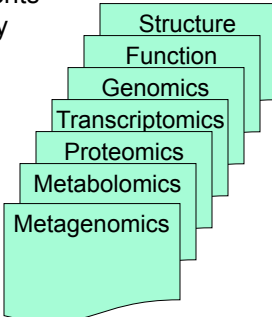
Reductionistic
Approach

Cell

Integrative
Approach

Components
Biology

Systems
Biology



Fall, 2015

GCBA 815

Sequenced Genomes listed by NCBI (As of 8/27/15)

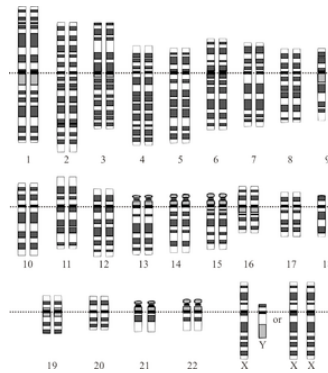
- Eukaryotes: 2,433
- Prokaryotes: 47,791
- Viruses: 4,903
- Plasmids: 6,156
- Organelles: 7,047

Fall, 2015

GCBA 815

Some Trivia about Human Genome

- # of chromosomes: 23 pairs, 24 distinct (22+X+Y)
- # of base pairs: 3 billion in haploid (diploid - 6×10^9 bp)
- Total length of stretched DNA: 2 meters
- # of Genes: ~ 25K-30K
- # of proteins: ~ 80K



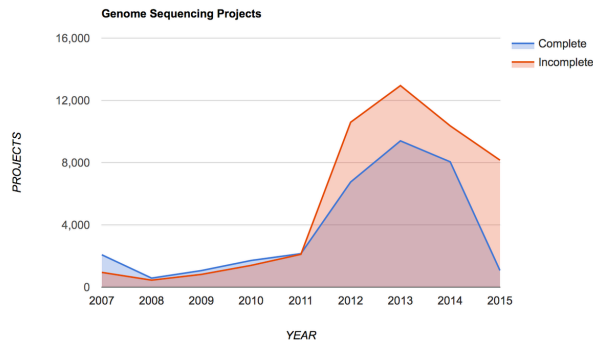
Fall, 2015

GCBA 815

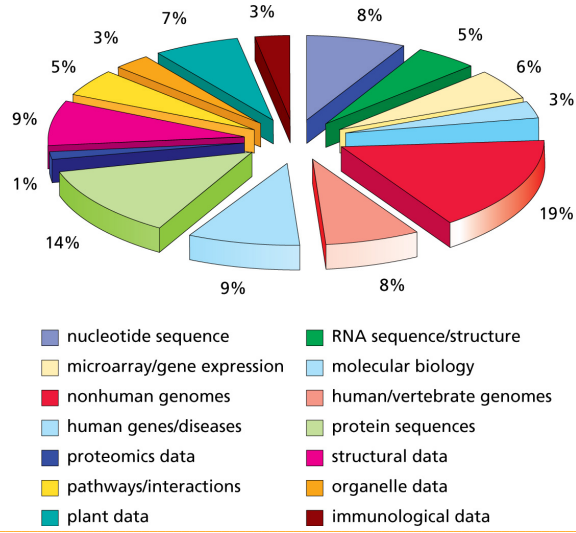
Chromosome	Length (cm)	Base pairs	Variations	Confirmed proteins	Putative proteins	Pseudogenes	miRNA	rRNA	snRNA	snoRNA	Misc ncRNA
1	8.5	249,250,821	4,401,091	2,012	31	1,130	134	66	221	145	106
2	8.3	243,199,373	4,607,702	1,203	50	948	115	40	161	117	93
3	6.7	198,022,430	3,894,345	1,040	25	719	99	29	138	87	77
4	6.5	191,154,276	3,673,892	718	39	698	92	24	120	56	71
5	6.2	180,915,260	3,436,667	849	24	676	83	25	106	61	68
6	5.8	171,115,067	3,360,890	1,002	39	731	81	26	111	73	67
7	5.4	159,138,663	3,045,992	866	34	803	90	24	90	76	70
8	5.0	146,364,022	2,890,692	659	39	568	80	28	86	52	42
9	4.8	141,213,431	2,581,827	785	15	714	69	19	66	51	55
10	4.6	135,534,747	2,609,802	745	18	500	64	32	87	56	56
11	4.6	135,006,516	2,607,254	1,258	48	775	63	24	74	76	53
12	4.5	133,851,895	2,482,194	1,003	47	582	72	27	106	62	69
13	3.9	115,169,878	1,814,242	318	8	323	42	16	45	34	36
14	3.6	107,349,540	1,712,799	601	50	472	92	10	65	97	46
15	3.5	102,531,392	1,577,346	562	43	473	78	13	63	136	39
16	3.1	90,354,753	1,747,136	805	65	429	52	32	53	58	34
17	2.8	81,195,210	1,491,841	1,158	44	300	61	15	80	71	46
18	2.7	78,077,248	1,448,602	268	20	59	32	13	51	36	25
19	2.0	59,128,983	1,171,356	1,399	26	181	110	13	29	31	15
20	2.1	63,025,520	1,206,753	533	13	213	57	15	46	37	34
21	1.6	48,129,895	787,784	225	8	150	16	5	21	19	8
22	1.7	51,304,566	745,778	431	21	308	31	5	23	23	23
X	5.3	155,270,560	2,174,952	815	23	780	128	22	85	64	52
Y	2.0	59,373,566	286,812	45	8	327	15	7	17	3	2
mtDNA	0.00054	16,569	929	13	0	0	0	2	0	0	22

Genome sequencing projects on GOLD

Genome Totals in GOLD (by year and status)



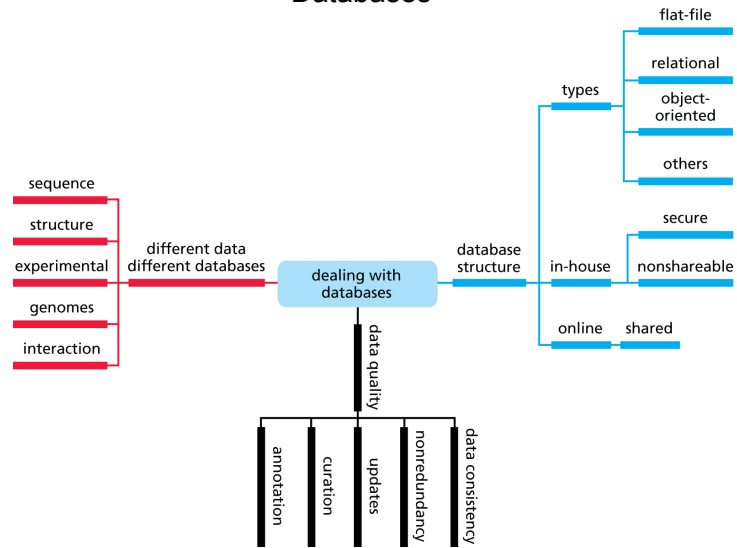
Different kinds of data in Bioinformatics



Fall, 2015

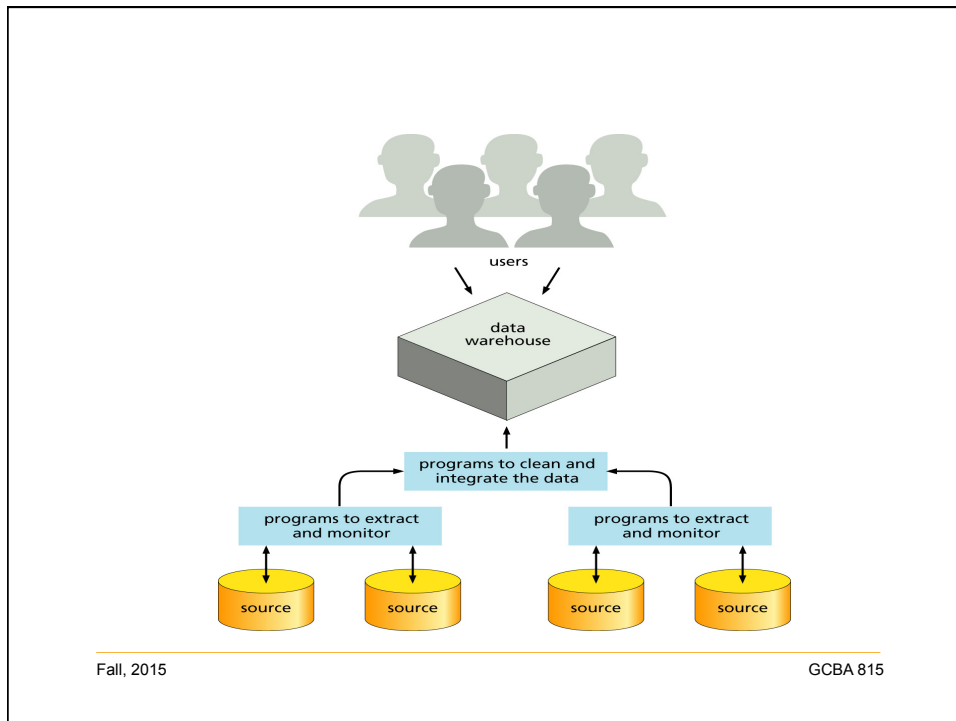
GCBA 815

Databases



Fall, 2015

GCBA 815



Sub-disciplines of Bioinformatics

- Data storage and retrieval (software, databases, web)
- Data curation and annotation (prediction tools)
- Data analysis (Genomics, proteomics, microarrays)
- Computational biology (Algorithms)
- Structural bioinformatics (3-D structures)
- Systems biology
 - Metabolic Pathways
 - Network biology
 - Genetic networks
 - Protein-protein interaction networks
 - Domain-domain interaction networks
- Synthetic Biology