

**BME 110L / BIOL 181L**

## **Computational Biology Tools**

**www.soe.ucsc.edu/classes/bme110/Winter09**

### **January 13:**

- Searching for similar sequences (to that of interest) in the databases (BLAST) - Introduction (Slides by Prof. Todd Lowe).  
[more advanced next week](#)
- A few pointers/reminders about HW assignments
- In-class demo: Using NCBI BLAST

Accompanying Reading (B4D): Chp 7; NCBI-BLAST tutorial

**BME 110L / BIOL 181L**

## **A question that came up a lot:**

### **What are COGs? (Clusters of orthologous groups)**

- the COG system attempts to provide an automated functional classification, i.e. assign a functional category to newly determined putative proteins based on sequence.
- this is done in a more reliable way than just “copying” the function of the best BLAST-hit with known function (more about orthologs/paralogs later today in the lecture)
- COG categories and assignments are often used to give a rough impression on what known biological processes are likely to be going on in the organism, and what proportion of genes is contributing to these
- uses a 1-Letter code to specify functional categories

## COG functional categories

Code	COGs	Domains	Description
<b>Information storage and processing</b>			
J	245	10,572	Translation, ribosomal structure and biogenesis
A	25	137	RNA processing and modification
K	231	11,271	Transcription
L	238	10,338	Replication, recombination and repair
B	19	228	Chromatin structure and dynamics
<b>Cellular processes and signaling</b>			
D	72	1,678	Cell cycle control, cell division, chromosome partitioning
Y	-	-	Nuclear structure
V	46	2,380	Defense mechanisms
T	152	7,683	Signal transduction mechanisms
M	188	7,858	Cell wall/membrane/envelope biogenesis
N	96	2,747	Cell motility
Z	12	128	Cytoskeleton
W	1	25	Extracellular structures
U	159	3,743	Intracellular trafficking, secretion, and vesicular transport
O	203	6,206	Posttranslational modification, protein turnover, chaperones
<b>Metabolism</b>			
C	258	9,830	Energy production and conversion
G	230	10,816	Carbohydrate transport and metabolism
E	270	14,939	Amino acid transport and metabolism
F	95	3,922	Nucleotide transport and metabolism
H	179	6,582	Coenzyme transport and metabolism
I	94	5,201	Lipid transport and metabolism
P	212	9,232	Inorganic ion transport and metabolism
Q	88	4,055	Secondary metabolites biosynthesis, transport and catabolism
<b>Poorly characterized</b>			
R	702	22,721	General function prediction only
S	1346	13,883	Function unknown

## BME 110L / BIOL 181L

### Other questions you ran into?

- 
- 
- 
-

## **BME 110L / BIOL 181L**

### **In a Nutshell: Absolutely Essential To Know**

#### **You should know at this stage...**

- what FASTA format is.
- what BLAST is for.
- your basic BLAST options (BLASTN, BLASTP, TBLASTN).
- that searching with protein sequence is more sensitive than searching with nucleic acid sequences (20 vs 4 types!)
- what an Expect value is, what it's for, what cut-off is reasonable
- what parameters are and that you have to keep track of them