"It is likely that unlikely things should happen."
-- Aristotle

"Probability is the very guide to life."
-- Marcus Tullius Cicero

Class Statistics

Total Students: 43 enrolled

Levels:

- Juniors: 20%
- Seniors: 31%
- Masters: 6%
- Doctoral: 43%

PC users: 1/3

Mac users: 2/3

Majors:

- MCDB: 32%
- Biochem: 20%
- Computer Sci: 9%
- IQ: 14%
- Engineer: 9%
- Ebio: 9%
- Other: 7%
  (Math, Iphy, Art history)

Class Statistics

Total Students: 43 enrolled

ZERO experience in:	+Limited
Biology: 4 (9%)	21%
Computer Science: 18 (42%)	56%
Statistics: 13 (30%)	58%

Desired Topics (most common)

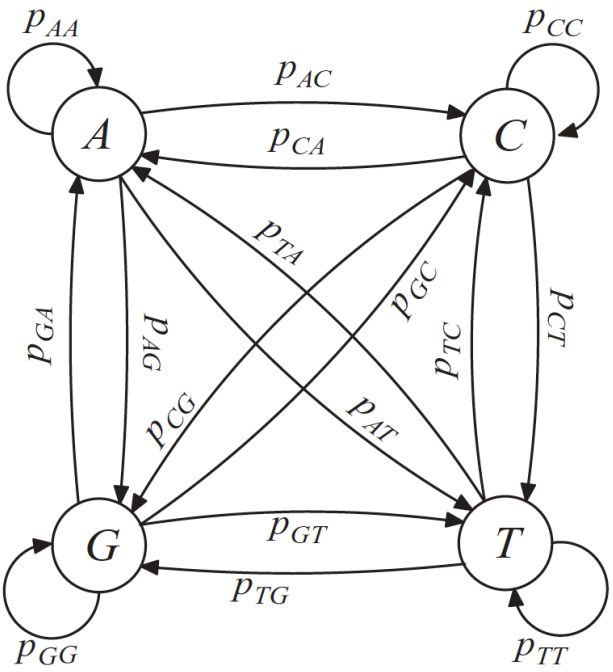
- Short Read Data Analysis Pipeline
- Genomics
- Statistics
- Understanding RNA-seq
- Proteomics

General Schedule now on Canvas

- First ~8 weeks are fundamental concepts in "Bioinformatics and Genomics": sequencing, alignment, HMMs, phylogenomics, annotation
- Next ~5 weeks is short read sequence analysis, including RNA-seq.
- Last ~2 weeks are proteomics.

Class will not meet on Friday (1/26)

- Instead, there are 3 short required videos on genome sequencing on Canvas.
- Quiz #1 will be available Friday 1/26 at 10:20 and close on Mon 1/29 at 9:20 am.



PROBABILITY

Because of the **stochastic** nature of genetics and evolution, we have to rely on the **theory of probability**.

Protein sequence: as a FASTA file

>gi|4504347|ref|NP_000549.1| **alpha 1 globin [Homo sapiens]**

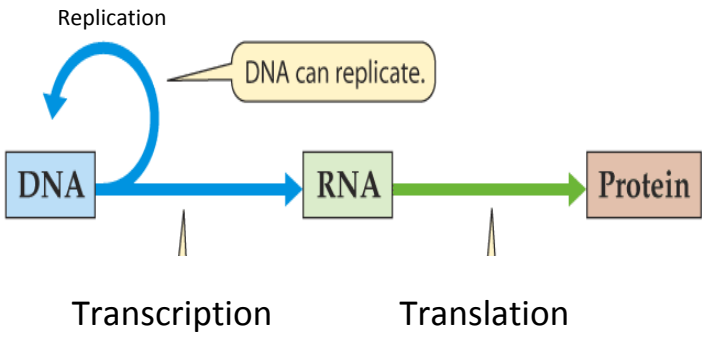
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAH
VDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR

DNA sequence: as a FASTA file

>gi|14456711|ref|NM_000558.3| **Homo sapiens hemoglobin, alpha** 1 (HBA1), mRNA

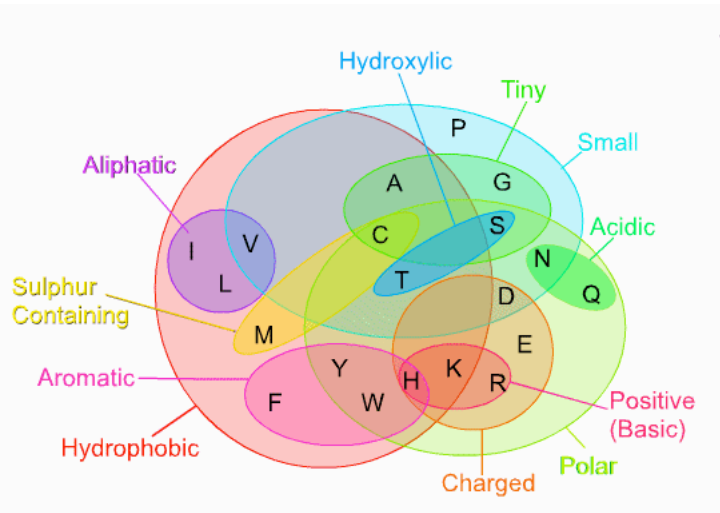
ACTCTTCTGGTCCCCACAGACTCAGAGAGAACCCACC**ATG**GTGCTGTCTCCTGCCGACAAGACCAACGTCAAGGCC
GCCTGGGGTAAGGTCGGCGCGCACGCTGGCGAGTATGGTGCGGAGGCCCTGGAGAGGATGTTCCTGTCCTTCCCCAC
CACCAAGACCTACTTCCCGCACTTCGACCTGAGCCACGGCTCTGCCCAGGTTAAGGGCCACGGCAAGAAGGTGGCCG
ACGCGCTGACCAACGCCGTGGCGCACGTGGACGACATGCCCAACGCGCTGTCCGCCCTGAGCGACCTGCACGCGCAC
AAGCTTCGGGTGGACCCGGTCAACTTCAAGCTCCTAAGCCACTGCCTGCTGGTGACCCTGGCCGCCCACCTCCCCGC
CGAGTTCACCCCTGCGGTGCACGCCTCCCTGGACAAGTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAATACC
GT**TAA**GCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTCCCCCCAGCCCCTCCTCCCCTTCCTGCACCC
GTACCCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGC

The Flow of Information



Second letter

First letter	U	C	A	G	Third letter
U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA Stop UAG Stop	UGU Cys UGC Cys UGA Stop UGG Trp	U C A G
C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg	U C A G
A	AUU Ile AUC Ile AUA Ile AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg	U C A G
G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gly GGC Gly GGA Gly GGG Gly	U C A G



Amino Acids

A alanine (ala)
R arginine (arg)
N asparagine (asn)
D aspartic acid (asp)
C cysteine (cys)
Q glutamine (gln)
E glutamic acid (glu)
G glycine (gly)
H histidine (his)
I isoleucine (ile)
L leucine (leu)
K lysine (lys)
M metioneine (met)
F phenyalanine (phe)
P proline (pro)
S serine (ser)
T threonine (thr)
W trytophan (trp)
Y tyrosine (tyr)

Amino Acid Percentages

Ala	(A)	7.81	Gln	(Q)	3.94	Leu	(L)	9.62	Ser	(S)	6.88
Arg	(R)	5.32	Glu	(E)	6.60	Lys	(K)	5.93	Thr	(T)	5.45
Asn	(N)	4.20	Gly	(G)	6.93	Met	(M)	2.37	Trp	(W)	1.15
Asp	(D)	5.30	His	(H)	2.28	Phe	(F)	4.01	Tyr	(Y)	3.07
Cys	(C)	1.56	Ile	(I)	5.91	Pro	(P)	4.84	Val	(V)	6.71

As determined by the SwissProt Database