Problem Set 0

This homework is intended to be a self-administered placement quiz, to help you (and me) determine if you have the background for the course or need to read additional material.

Collaboration is not allowed on this homework. You may consult only with me and the TA. You may refer to textbooks and other written materials, but not to solution sets from previous years.

Due September 1st

1. This question asks you to compare properties of typical animal, plant and bacterial cells. It is not necessary to consider unusual species with properties that differ from the cellular organization of commonly studied organisms.

   (a) **Animal cells:** The left most column in the table below gives a list of cellular features, i.e. a molecule, molecular complex or cellular process. The top row lists various cellular compartments. In each row, put an 'X' under each compartment in which the corresponding cellular feature occurs in *animal cells*. For any given row, zero, one, or more than one 'X' may be required.

	Chloroplast	Cytoplasm	Mitochondrion	Nucleus
Chromosomal crossover				**X**
Circular chromsomes			**X**	
Diploid chromosomes				**X**
DNA			**X**	**X**
mRNA		**X**	**X**	**X**
Ribosomes		**X**	**X**	
Splicing				**X***
Transcription			**X**	**X**
Translation		**X**	**X**	
tRNA		**X**	**X**	

* Organelle mRNA may be intronic(No points deducted for this).

(b) **Bacterial cells:** The left most column in the table below gives a list of cellular features, i.e. a molecule, complex or process. The top row lists various cellular compartments. In each row, put an 'X' under each compartment in which the corresponding cellular feature occurs in *bacterial cells.* For any given row, zero, one, or more than one 'X' may be required.

	Chloroplast	Cytoplasm	Mitochondrion	Nucleus
Chromosomal crossover				
Circular chromsomes		**X**		
Diploid chromosomes				
DNA		**X**		
mRNA		**X**		
Ribosomes		**X**		
Splicing		**X***		
Transcription		**X**		
Translation		**X**		
tRNA		**X**		

* Some non coding RNAs are spliced(No points deducted for this).

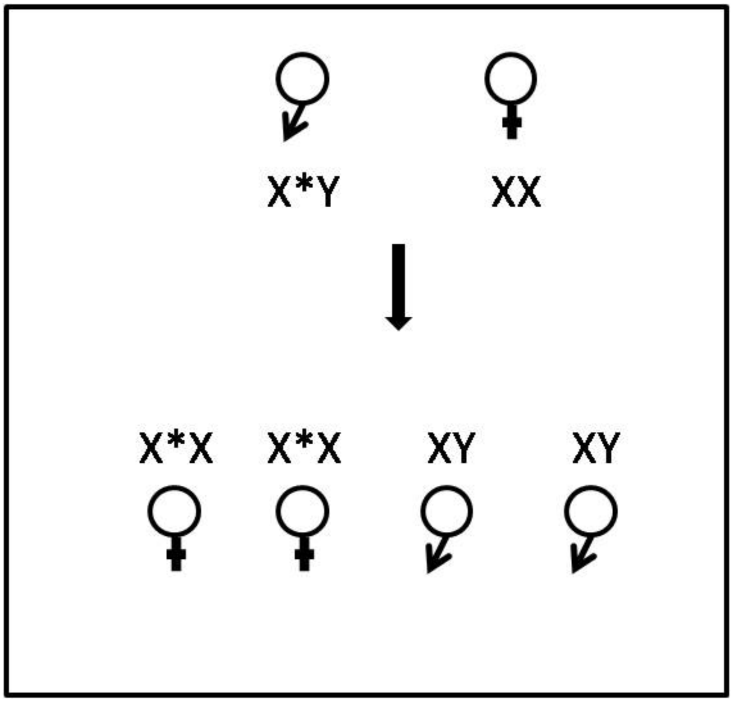
(c) **Plant cells:** The left most column in the table below gives a list of cellular features, i.e. a molecule, complex or process. The top row lists various cellular compartments. In each row, put an 'X' under each compartment in which the corresponding cellular feature occurs in *plant cells*. For any given row, zero, one, or more than one 'X' may be required.

	Chloroplast	Cytoplasm	Mitochondrion	Nucleus
Chromosomal crossover				**X**
Circular chromsomes	**X**		**X**	
Diploid chromosomes				**X**
DNA	**X**		**X**	**X**
mRNA	**X**	**X**	**X**	**X**
Ribosomes	**X**	**X**	**X**	
Splicing				**X***
Transcription	**X**		**X**	**X**
Translation	**X**	**X**	**X**	
tRNA	**X**	**X**	**X**	

* Organelle mRNA may be intronic(No points deducted for this).

2. An X-linked dominant allele causes *vitamin D resistant rickets*, a form of rickets due to low serum phosphorus in humans. A man with vitamin D resistant rickets marries a normal woman. What proportion of their sons will have rickets? What proportion of their daughters will have rickets? (Use the symbol X^R for the mutant and X for the normal or *wild type* allele.)

The man passes on his Y chromosome to his sons and the X^R *to his daughters. So no son has rickets(0%) but all daughters will have(100%).*
In the figure, ***X**** *denotes the dominant allele and X the normal allele.*



3. To solve this problem, you will need a table of the genetic code and the following table giving the physico-chemical properties of the 20 amino acids:

small	**hydrophobic**		**polar**	**acidic**	**basic**
Gly	Val	Phe	Asn	Asp	Lys
Ala	Cys	Tyr	Gln	Glu	Arg
Ser	Ile	Met	His		
Thr	Leu	Trp			
	Pro				

(a) Consider the tenth codon in the folloowing (very short) gene:

ATG GCA AGA AGC GCA ACA ACG GCG TGT [AAG] AGA TAA

What amino acid does it encode and what physico-chemical class does it belong to?

AA : Lys
Class : Basic

(b) Write down all codons that can result from a single base-pair replacement in the tenth codon.

AAT	**TAG**	**ACG**
AAC	**CAG**	**ATG**
AAA	**GAG**	**AGG**

(c) What is the probability that a single base change in this codon results in a substitution of the associated amino acid for one in the same class?

Lys : AAA
Arg : AGG

Probability = 2/9 = 0.22

(d) What is the probability that a single base change in this codon leaves the amino acid unchanged?

Lys : AAA
Probability = 1/9 = 0.11

(e) What will be the impact on the protein sequence if the first adenine in the codon is replaced by a thymine?

A replacement of the first A to T leads to a termination of protein translation as TAG encodes the STOP codon.

4. Phenylketonuria (PKU) is a recessive disease caused by mutations in the phenylalanine hydroxylase gene on chromosome 12 in humans. In PKU, phenylalanine, which is toxic to the brain, is not broken down and accumulates in the blood. Untreated individuals with PKU show progressive developmental delay in the first year of life, mental retardation, seizures, autistic-like behavior and a peculiar body odor. Approximately 1 individual in 50 is a carrier.

   (a) A phenotypically normal man whose father had PKU marries a phenotypically normal woman from outside the family, and the couple consider having a child.

   What are the possible genotypes of the husband? Of the husband's father? Of the wife? (Use the symbol "a" for the mutant and "A" for the normal or *wild type* allele.)

   Husband: Aa

   Wife: Aa

   Father: aa

   (b) Given this information and the frequency of carriers in the population, what is the chance that the couple's first child will have PKU?

   P(a from father) $= \frac{1}{2}$
   P(a from mother) $=$ *P(mother is Aa)* $* \frac{1}{2} = \frac{1}{50} * \frac{1}{2}$
   P(PKU) $=$ *P(a from father)* $*$ *P(a from mother)* $= \frac{1}{2} * \frac{1}{50} * \frac{1}{2} = \frac{1}{200}$

   (c) If the first child does have PKU, what is the probability that the second child will be normal?

$$1 - \frac{1}{2} * \frac{1}{2} = \frac{3}{4}$$

   Since the first child has PKU, the wife's genotype must be Aa.

   (d) The incidence of PKU is roughly 1 in 16,000 individuals worldwide. Diagnostic tests in infants have a 90% false positive rate. The probability of a false negative approaches zero by the third day of life. Assuming that the false negative rate is zero, what is the probability that a randomly selected infant who tests positive, actually has the disease? (Hint: use Bayes' theorem)

   Let E_D *be the event that the patient has PKU,* E_H *be the event that the patient is healthy,* E_P *be the event that the patient's test is positive*

$$\begin{aligned} P(E_D|E_P) &= \frac{P(E_D) * P(E_P|E_D)}{P(E_D) * P(E_P|E_D) + P(E_H) * P(E_P|E_H)} \\ &= \frac{\frac{1}{16{,}000} * 1}{\frac{1}{16{,}000} * 1 + \frac{15{,}999}{16{,}000} * 0.9} \\ &= 6.94 * 10^{-5} \end{aligned}$$

5. A clique is an undirected graph $G = (V, E)$ in which every pair of vertices is connected by an edge. Suppose $|V| = n$.

   (a) Give the number of edges in G in terms of n.

   $\frac{n(n-1)}{2}$

   (b) What is the degree of each vertex, v?

   n-1

   (c) Let T_b be the spanning tree obtained from a breadth first search of G. Give the number of edges in T_b in terms of n.

   n - 1

   (d) The diameter of a tree is defined to be $\max_{u,v \in V}\{\text{the shortest path from } u \text{ to } v\}$. What is the diameter of T_b?

   2

   (e) Let T_d be the spanning tree obtained from a depth first search of G. Give the number of edges in T_d in terms of n.

   n - 1

   (f) What is the diameter of T_d?

   n - 1

6. Assume that the trees in all of the following problems are *rooted*. A complete binary tree is a tree in which every node has either zero or two children. Let T be a complete, rooted binary tree with L leaves.

   (a) Give the number of edges in T in terms of L.

   2L - 2

   (b) Give the number of nodes in T in terms of L.

   2L - 1

   (c) What is the maximum possible depth of a complete, rooted binary tree with L leaves? What is the minimum possible depth of a complete, rooted binary tree with L leaves? Give your answer in terms of L.

   Max depth : L - 1
   Min depth : $\lceil \log_2 L \rceil$

   (d) Which of the following data structures would you use to keep track of the next node to visit in implementing a breadth-first-search? A depth-first-search?

   i. A sparse matrix
   ii. A stack
   iii. A hash table
   iv. A queue
   v. A heap

   BFS : Queue
   DFS : Stack

   (e) In a connected graph, an articulation point is a node with the property that, if the node is removed, the two or more disconnected subgraphs will result. How many articulation points are there in a clique with N nodes? How many articulation points in a tree with L leaves?

   Clique : 0
   Tree : L - 1(UNROOTED), L - 2(ROOTED)

7. Algorithms X and Y have worst case running times no greater than $150N \cdot \sqrt{N}$ and N^2, respectively. Which algorithm has better asymptotic running time? For which values of N would you choose algorithm X over algorithm Y? (You may give an algebraic or graphical answer.)

   X has a better asymptotic running time as X is $O(N^3/2)$ and Y is $O(N^2)$.

   We choose algorithm X over algorithm Y when $150N \cdot \sqrt{N} < N^2$
   i.e. when $N > 22500$