Name: _____ SID#: _____

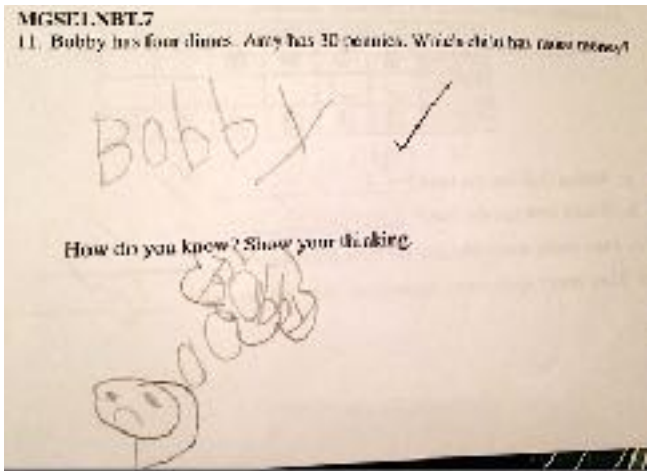
2nd Midterm **MCB 140 Fall 2015** **5 November 2015**

NAME (Please print):_____
STUDENT ID #:_____

You have 2 hours for this 150 point exam.
As advertised in lecture, there is a three point extra credit question.

Please print your name and ID# on **each** page of the exam; **a point will be deducted if your name is not on every page**. There are 7 pages in total (including this cover sheet).

Please turn in all 7 pages of this exam; note that only the front of each page will be graded. Please write in pen, not in pencil. Thank you!



Do not write below this line.

Page 2 (40) _____

Page 3 (15) _____

Page 4 (25) _____

Page 5 (25) _____

Page 6 (30) _____

Page 7 (10 + extra credit) _____

TOTAL _____ / 150

Name: _____ SID#: _____

Name: _____ SID#: _____

Problem 1 (5 points). What is the typical size of a human chromosome (in megabases)?

100 megabases is a typical human chromosome size. You can either remember this from lecture or divide the genome size (3 x 10^9) by chromosome number (23). Anything in the 50-200 Mbp range is full credit.

Problem 2 (5 points). Haldane measured the rate at which X-linked mutations

Note: this question was accidentally truncated. Full credit (5 points) given to all.

Problem 3 (5 points). What is typical recombination rate (cM/Mb) in humans?

1 cM/Mb is the typical recombination rate. You can either remember this or note that there is ~1 crossover per chromosome, or 100 cM / 100 Mbp ~ 1 cM/Mbp.

Problem 4 (5 points). Give a simple definition of *narrow sense heritability* of a trait.

As described in class, narrow sense heritability is the slope of the (best fit) line of the progeny's trait vs. the midparent trait (i.e., average of the parental straits)

Problem 5 (5 points). A mutation introduces an extra base pair within coding sequence near the beginning of a gene. Explain in a sentence or two why you expect this to be a loss of function mutation.

This is a frameshift mutation. It shifts all subsequent codons and is likely to produce a stop codon in the new reading frame. Since its at the beginning of the gene the protein produce will be severely truncated, resulting in a loss of function. (Full credit, both ideas must be present)

Problem 6. Sequencing (15 points). Sanger sequencing takes advantage of fundamental features of molecular biological reagents. Explain the importance of **three** of the following in a sentence or two. (5 points each).

Dideoxy nucleotides.
Fluorescent dyes.
DNA polymerase.
Cloning vector.
Gel electrophoresis.

DNA whose sequence is unknown can be easily inserted into a **cloning vector,** which allows that sequence to be amplified indefinitely within bacteria**. Dideoxy nucleotides** can be added to a newly synthesized DNA strand but cannot be extended further by **DNA polymerase** (they are "terminators"). If the dideoxy nucleotide is attached to a **fluorescently dye,** with each nucleotide having a different color, then strands will be labeled according to the identify of their

Name: _____ SID#: _____

final nucleotide. Gel **electrophoresis** allows DNA of different length to be separated. We can then read off the nucleotide as a particular distance from the primer site based on its color. Voila!

Problem 7. Three point test cross (40 points). To test your understanding of the three point test cross we will use three linked recessive seed traits in maize.

S is plump wild type allele; s is the recessive shrunken allele
W is wild type kernel coating allele; w is the recessive waxy allele.
C is white wild type allele; c is the recessive red (colored) allele

We start with two inbred lines L1 and L2. Line L1 has genotype ssWWCC (sWC/sWC) and Line L2 has genotype SSwwcc (Swc/Swc). (Caution: we do not yet know the order of the genes on the chromosome; this is for you to determine below!)

(7a, 5 points) What are the **phenotypes** of Line L1 and Line L2?

Noting the stated dominant/recessive relationships between alleles,
Line L1 is shrunken, wild kernel, white.
Line L2 is plump, waxy, red.

(7b, 5 points) We next cross Line L1 and Line L2 to form a trihybrid T. What is the **genotype** and **phenotype** of the trihybrid?

Since the trihybrid L1 x L2 is heterozygous at every locus, it has the dominant traits: plump, wild type kernel, white. The genotype is sWC/Swc

(7c, 5 points) To perform a testcross, we cross the trihybrid with the triple recessive sswwcc (swc/swc). Explain in a sentence or two the advantage of using a triple recessive in a testcross.

Because one of the parents in the testcross is recessive at every locus, the phenotype of the progeny is completely determined by the gamete that is provided by the other parent. So the haplotype of the gamete can be read off directly by just looking at the plant.

Seed trait	Gamete from trihybrid parent	Count	Percentage
shrunken, normal, red	sWC	2,777	39.7%
plump, waxy, white	Swc	2,708	38.7%
shrunken, waxy, red	swC	643	9.2%
plump, normal, white	SWc	626	8.9%
shrunken, normal, white	sWc	123	1.8%
plump, waxy, red	SwC	116	1.7%

Name: _____ SID#: _____

plump, normal, red	SWC	4	<0.1%
shrunken, waxy, white	swc	3	<0.1%
Total number of progeny:		7,000	100%

(7d, 5 points). Which line or lines in the table represent unrecombined gametes (parental haplotypes)?

Since the parent is he unrecombined gametes are represented by the top two lines of the table: sWC and Swc.

(7e, 5 points). From the table, what is the frequency of observed recombination between the s and w loci? w and c? s and c? Call these three numbers rec(s,w), rec(w,c), and rec(s,c).

rec(s,w) = gametes that have the opposite phase from the parents at these loci = SWx and swx, where x is either C or c. Note that we don't care what is happening at locus c. = 9.2% + 8.9% + (two small numbers that add to 0.1%) = 18.2% (full credit if you ignored the infrequent SWC and swc and got 18.1%)

rec(w,c) = gametes xWc or xwC = 9.2% + 8.9% + 1.8% + 1.7% = 22.6%

rec(s,c) = gametes sxc or SxC = 1.8% + 1.7% + (two small number that add to 0.1%) = 3.6% (full credit if you ignored the infrequent SWC and swc and got 3.5%).

(7f, 5 points). From (e), how are the loci ordered along the chromosome? If you were not able to complete (e), describe how you would have done it given rec(s,w), rec(w,c), and rec(s,c).

s and c are closest (lowest recombination frequency).

Of these two, w is closest to s, and farthest from c.

So the order is wsc or csw (either answer is correct).

(7g, 5 points). Explain in a sentence or two why are the bottom two entries so small?

These are gametes corresponding to double crossovers. Since a single crossover is infrequent (~3-22% as computed above) the double crossover is rare.

(7h, 5 points). Explain in a sentence or two why map distances tend to be underestimated when genes are further apart?

For nearby loci, the map distance is very close to the recombination frequency. But as we consider loci that are farther apart, there is an increased chance that a double crossover will occur

Name: _____ SID#: _____

between the two loci. But since these cannot be detected (unless there are markers in between), we don't detect these double crossovers and underestimate the map distance.

Problem 8. *A variant of* Beadle's maize scenario (15 points). Beadle envisioned the domestication of maize from wild teosinte as involving homozygosing a small number of *recessive* "domestication" alleles in maize. That is, at some loci A, B, C, … , the genotype of teosinte is A^TA^T, B^TB^T, C^TC^T, … and the genotype of maize is A^MA^M, B^MB^M, C^MC^M, … We will assume that the loci are unlinked (*e.g.*, on different chromosomes).

(8a, 5 points) In this simple model, what is the **phenotype** of the F1 hybrid of teosinte and maize?

An F1 hybrid has one allele from maize and one from teosinte. The maize alleles are assumed to be recessive. So the phenotype is teosinte.

(8b, 5 points) Assume that there are three loci A, B, and C. Write out <u>all</u> of the possible genotypes arising from a **backcross** F1 x M. (**Warning: this is NOT the Beadle F2 cross that we discussed in class**.)

F1: A^TA^M, B^TB^M, C^TC^M. We are backcrossing to maize to the maize gamete is always $A^MB^MC^M$

There are eight possible progeny of F1 x M =

A^TA^M, B^TB^M, C^TC^M
A^TA^M, B^MB^M, C^TC^M
A^MA^M, B^MB^M, C^TC^M
A^TA^M, B^MB^M, C^MC^M
A^MA^M, B^TB^M, C^TC^M
A^TA^M, B^TB^M, C^MC^M
A^MA^M, B^TB^M, C^MC^M
A^MA^M, B^MB^M, C^MC^M

(8c, 5 points) Now generalize from n=3 loci in part (b) to any n. If we grow out a thousand plants from this backcross and find that ~3% of the progeny are maize-like, can you estimate n? You can provide a number or clearly describe a calculation.

For n=3, one out of eight (12.5%) will be maize-like (the last one listed above). In general the fraction is $(½)^n$. For n=4, its one out of 16 (6.25%). For n=5 it is one out of 32 (3.125%). So we infer that there are ~5 loci that are responsible for the maize-like phenotype in this model.

Problem 9. Broad-sense heritability (10 points). Two **inbred lines** of beans are intercrossed. In the F1 hybrid, the variance in bean weight is measured at 1.5 gm^2. The F1 is selfed to produce an F2 generation. In the F2 population, the variance in bean weight is 6.1 gm^2.

Estimate the **broad sense heritability** of bean weight in the F2 population. Be sure to explain the definition of broad sense heritability and your logic in estimating it.

Name: _____ SID#: _____

The F1 population is genetically identical (it is a cross between two inbred lines). So all of its variance is "environmental": Var(E) = 1.5 gm^2. In the F2 population, we have both genetic and environmental variance Var(P) = Var(G) + Var(E) = 6.1 gm^2. We can then solve for the genetic variance: Var(G) = Var(P) – Var(E) = 6.1 – 1.5 = 4.6 gm^2
Broad sense heritability = Var(G)/Var(P) = 4.6/6.1 (=0.75).

Problem 10. Allele and genotype frequencies (15 points). You sample 100 individuals and genotype them at an autosomal biallelic SNP with alleles A and G. In your sample there are 50 individuals that are AA homozygotes, 20 individuals that are AG heterozygotes, and 30 individuals that are GG homozygotes.

(10a. 5 points) What are the allele frequencies of the A allele (call it *p*) and the G allele (*q*)?

p = (number of A alleles)/(number of total alleles) = (2 x 50 + 1 x 20)/200 = 0.6
q = (number of G alleles)/(number of total alleles) = (1 x 20 + 2 x 30)/200 = 0.4
Or you could compute one of them and infer the other from p+q=1.

(10b. 5 points) What would the frequency of AA homozygotes be if the population had been in Hardy-Weinberg equilibrium? Is the population in Hardy-Weinberg equilibrium?

According to the Hardy-Weinberg formula, if this population were mating randomly, the frequency of the AA genotype is $p^2 = 0.36$. The population is NOT in Hardy-Weinberg equilibrium.

(10c. 5 points) If the population you sampled mated randomly to produce the next generation, would you expect the number of heterozygotes to increase, or decrease?

If the population mates randomly the number of heterozygotes will be 2pq = 0.48. It will go up. Full credit if you gave some other heuristic explanation for why it should increase.

11. Human mitochondrial genetics (20 points).

(11a .5 points.) True or false: only females have mitochondrial genomes.

False! All human cells have mitochondria, and each comes with a genome. Males just don't transmit them to the next generation

(11b. 5 points.) Describe in a sentence or two the evidence that the mitochondrial ancestor of humans lived in Africa.

The greatest mitochondrial diversity is in Africa. (This is sufficient for full credit.) Also, the mitochondrial diversity outside Africa is largely a subset of that found in Africa.

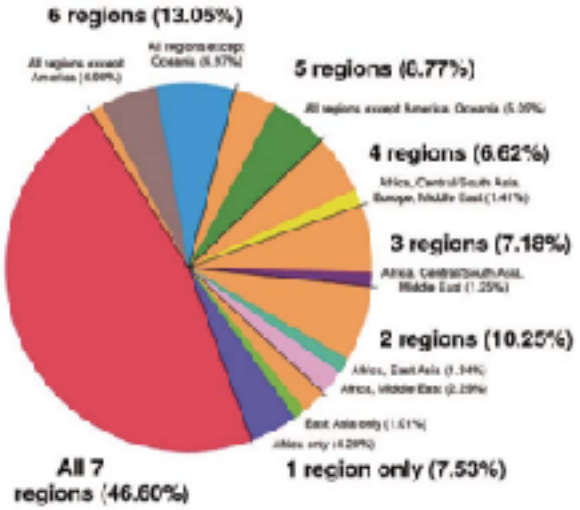
Name: _____ SID#: _____

(11c. 5 points.) Explain why there is less mitochondrial variation in the Americas than in Eurasia.

The Americas were populated by migration from northeast Asia. Due to a founder effect, the amount of mitochondrial variation in the Americas was reduced.

Name: _____ SID#: _____

12. Variation in a population (10 points). In lecture we looked at data on thousands of microsatellite loci. It was found (as shown in the pie chart below) that nearly half of all alleles are found in all continental regions, with only a few percent unique to a given continent. In a few sentences, state whether you think the alleles found across all regions are typically older or younger than those found in a single continent. You may use any information you remember from the lectures as well as population genetic logic.



Alleles found across all continents must be older, since their birth (by mutation) must have predated the migration of humans to populate other continents. Conversely those alleles with a more limited geographic distribution (say, only one continent) are more likely to have arisen by recent mutation in that region.

13. Mutation-selection balance (5 points). Consider a gene whose loss of function causes a **recessive lethal disease**. Assume that the mutation rate from "wild type" to "loss of function" is $\mu = 10^{-6}$.

If the locus is in mutation-selection balance, what is the frequency of **carriers** of the disease alleles? (Note here you are asked for carriers, not the allele frequency.)

Allele frequency q is square root of μ , or 10^{-3}.

Frequency of **carriers** in Hardy-Weinberg equilibrium is $2pq \sim 2 \times 10^{-3} = 2/1000$ or 0.2%

3 points off if you gave the allele frequency rather than the carrier frequency.

14. Extra credit (3 points). Name one band that Grace Slick performed with. Anything close gets credit.

Name: _____ SID#: _____

I listed these in class and on bcourses. Jefferson Airplane, then Jefferson Starship, then Starship.