- **Put your name and your GSI's name on ALL 9 pages. The pages will separated for grading.**
- **The exam is closed book: no papers, notes, or books may be used. No calculators, computers, telephones, mp3 players or other electronic devices may be used at any time during the exam.**
- **You can leave your answers as fractions and sums.**
- **Write all answers in the space provided. You can use the backs of the pages for your own notes.**
- **Be sure to answer all parts of all 21 questions.**
- **The numbers in parentheses are the points for each part. The total is 150.**

1. (6) What are two differences between SNPs and SSRs that affect their use as genetic markers? [Hint: The question is not what are SNPs and SSRs.]

(i) SNPs are have only two alleles while SSRs typically have many alleles. (ii) SNPs are more abundant in the genome than SSRs. It would also be correct to say that SNPs are easier to genotype using chips while SSRs require PCR amplification.

2. a. (3) What is a genetic map of a chromosome?

A list of gene (or locus) positions and the recombination distance between them.

b. (3) What is a physical map of a chromosome?

A list of gene positions and the physical distances (or numbers of base pairs separating them).

c. (4) What is likely to be different and what is likely to be the same in physical and genetic maps of the same chromosome?

Relative distances are likely to be different but gene order is likely to be the same.

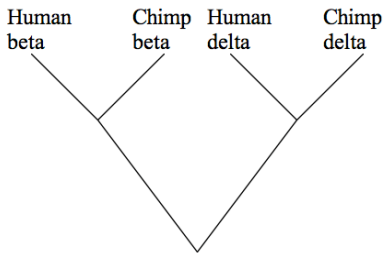
3. (3) Suppose you found a species which used an unusual genetic code in which there are 4 instead of 3 stop codons. In this species what would be the average number of nucleotides between two stop codons in randomly assembled DNA when read in a single reading frame?

64/4 =16 codons = 48 bases. 45 bases is OK also.

4. (3) Roughly 50% of the human genome is repeat sequence. What elements make up the majority of the repeat sequence?

transposons

5. (8) The two loci coding for beta-globin and delta-globin are in the same gene region on chromosome 11 in humans. Draw a gene tree that describes the ancestry of these two loci on one chromosome from a human and one chromosome from a chimpanzees. Indicate which loci are paralogous and which are orthologous.



Human beta and chimp beta are orthologous. Human delta and chimp delta are orthologous. Human beta and human delta are paralogous. Chimp beta and chimp delta are paralogous. It is OK but not necessary to say Human delta and chimp beta, and human beta and chimp delta are paralogous, but take points off if they say they are orthologous. No points for saying they are homologous.

6. a. (5) What is recurrent mutation?

Recurrent mutation is a specific nucleotide change that has occurred more than once.

b. (5) What information is used as evidence of recurrent mutation?

The observation that there are two (or more) haplotype backgrounds associated with the same mutation indicate that there was recurrent mutation.

7. (6) What are two observations that indicate there has been balancing selection affecting genes in the MHC complex in humans and related species?

(i) MHC (or HLA) loci in humans have many alleles, more than are found at other loci.
(ii) At some HLA loci (DQB1) there is evidence of trans-species polymorphism.

8. (4) Assume you have genotyped a parent-offspring trio at three loci and found the following data. [Assume there is no recombination.]

Mother	Father	Offspring
AABbcc	AabbCc	AaBbCc

What is the haplotype phase of the offspring?

ABc/abC. The phase of the mother has to be ABc/Abc.

9. (5) A microsatellite locus in a randomly mating population has five alleles, $A_1 \ldots A_5$ with frequencies 0.4, 0.35, 0.1, 0.08 and 0.07 in that order. What is the fraction of the population that is heterozygous for A_5?

2x0.07x0.93. 2x0.07x(0.4+0.35+0.1+0.08) is also OK

10. (8) What are the two observations that indicate there was positive selection on lysozymes in langurs?

(i) There are many more amino acid substitutions on the lineage leading to langurs than on the lineage leading to baboons. (ii) Several (5) of the substitutions in langurs are the same as found in cows.

11. (4) Approximately 1 in 10,000 males in the US has hemophilia caused by a mutation of the Factor VIII gene on the X chromosome. This mutation is recessive in females. What fraction of the female US population will have hemophilia if the genotype frequencies are in their Hardy-Weinberg proportions?

$(1/10000)^2=10^{-8}$ Either answer is correct.

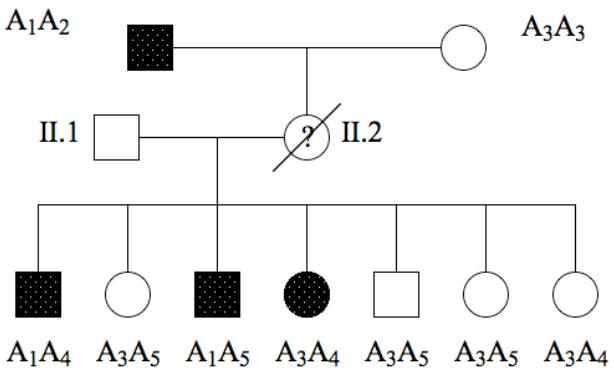
12. (4) A population of diploid self-fertile plants on a new volcanic island is established by a single seed from a source population on the nearby mainland. You have genotyped two unlinked loci in the source population and found the following genotype frequencies:

AA 0.5, Aa 0.2, aa 0.3
BB 0.4, Bb 0.2, bb 0.4

What is the chance that the island population is fixed for A and b? Assume that the seed was chosen randomly from the source?

0.5x0.4=0.20 = 20% Anything involving Hardy-Weinberg is wrong.

13. In the pedigree below, $A_1 \ldots A_5$ are alleles of a single locus. The filled symbols indicate affected individuals and the question mark indicates that the affected status is unknown.



a. (4) What are the genotypes of II.1 and II.2 at the A locus? [Assume there is no mutation.]

II.1 is A_4A_5; II.2 is A_1A_3

b. (8) Assume that the affected status of an individual is determined by a dominant allele with complete penetrance. Find the odds ratio in favor of the hypothesis that the recombination rate between the A locus and the disease locus is 1/4.

II.2 is affected and the haplotype phase is A_1D/A_3d, where D is the allele that causes the disease. There are 6 non-recombinants and 1 recombinant (the affected female). Therefore the odds ratio is

$$\frac{(1-1/4)^6(1/4)}{(1/2)^7}$$

14. A dominant mutation in *Drosophila* called *Delta* causes abnormal wing morphology in *Delta*/+ heterozygotes. Homozygosity for this mutation (*Delta*/*Delta*) is lethal. In a population of 3000 flies, 1500 had normal wings and 1500 had abnormal wings.
a. (3) What is the frequency of *Delta* in this population?

¼

b. (6) Using the allele frequency calculated in part a, how many total zygotes must be produced by this population in order for you to count 3000 viable adults in the next generation?

(16/15)x3000. 1/16 of the zygotes will die because they are homozygous for *Delta*.

c. (6) Assume there is random mating, no migration, and no mutation, and ignore the effects of genetic drift. What are the expected numbers of the two adult phenotypes (normal and abnormal wings) if 3000 viable offspring of the population in part a are counted?

normal wings (9/16)(16/15)3000, abnormal wings (6/16)(16/15)3000.

15. In non-European populations, cystic fibrosis (CF) is a rare condition. Assume alleles of CFTR that cause CF are recessive and fully penetrant, and assume that genotype frequencies in newborns are in their Hardy-Weinberg frequencies and the frequency of CF among newborns is 1/10,000.
a. (5) What is the fraction of newborns who are heterozygous carriers of CF alleles?

$p=\sqrt{(1/10{,}000)}=1/100$, Therefore $2pq=2\text{x}0.01\text{x}0.99\approx0.02$

b. (5) If the mutation rate from normal to CF alleles is 10^{-6} per generation and the population is at an equilibrium under mutation-selection balance, what is the selection coefficient against newborns with cystic fibrosis in this (hypothetical) population?

The formula is $p=\sqrt{(\mu/s)}$ with $p=10^{-2}$ and $\mu=10^{-6}$. Therefore $\sqrt{s}=1/10$ or $s=1/100$.

16. (6) In a study of a 10,000 base region of gorilla chromosomes, the average pairwise sequence difference is 25 . Assume that the neutral mutation rate is $2x10^{-9}$ and estimate the population size of gorillas that is consistent with these data.

The formula is $\pi=4N\mu L$. Therefore, $N=25/(4x2x10^{-9}x10{,}000)=25/8 \times 10^5$.

17. The following table shows the polymorphic sites found in a sample of Y-chromosomes. Assume sample IV is the outgroup.

	SNP										
	1	2	3	4	5	6	7	8	9	10	11
I	C	T	A	T	A	T	C	G	T	C	T
II	C	T	A	A	G	T	C	G	T	C	G
III	T	T	A	T	G	T	C	G	T	C	T
IV	C	G	G	T	G	G	A	A	G	A	T

a. (5) Which tree is favored based on the pairwise differences in sequence?
The distances are

	II	III	IV
I	3	2	8
II		3	9
III			8

although they do not need to compute all the distances. The I-III tree is favored.

b. (5) Which tree is favored based on parsimony?

None of them is favored by parsimony. (Or all of them are equally favored by parsimony.)

18. (5) Describe one property of the gene tree of mtDNA that would have supported the local continuity theory of the origin of modern humans?

Either (i) The time of the MRCA would be at least 1.5 million years, the time when *H. erectus* dispersed from Africa. or (ii) The gene trees for mtDNAs from Europe, Africa and Asia would form monophyletic groups. No extra credit for giving both answers.

19. (5) There is genetic evidence for a bottleneck after the ancestors of humans separated from the ancestors of chimpanzees. What is that evidence?

The per site difference between two sequences in humans (π) is about 0.001, indicating an population size of slightly more than 10,000 individuals (11,250). This population size is much smaller than the current size of the human populations. And π for humans is less than half of π for chimpanzees.

20. (4) Assume that phenylketonuria (PKU) is caused by completely penetrant recessive mutations at the PAH locus. What is the recurrence risk to full siblings of a proband if neither parent has PKU?

1/4

21. In humans, the locus ALDH codes for an enzyme important for the metabolism of alcohol. Individuals homozygous for null mutations at this locus cannot metabolize alcohol. In some human populations, null alleles ALDH have reached a frequency of 30%.

a. (6) What evidence would you look for to support the claim that null alleles at ALDH rose to high frequency because of positive selection?

If null alleles were associated with a long haplotype that was not associated with other alleles, that would indicate positive selection.

b. (4) What would be another explanation for the high frequency of null alleles at ALDH in a population?

A founder effect (or genetic drift) could also result in a high frequency of a null allele.