

1. (15 points) Cytosine deamination produces what base?

3 points - **uracil**

What causes cytosine deamination (X-rays? uv light? alkylating agents? something else?)?

3 points – **This reaction is spontaneous (water-catalyzed).** Drawing the mechanism for this acid/base reaction was good, but not required for full credit. Alkylation can also cause cytosine deamination, so this answer was accepted.

How are deaminated cytosines in DNA repaired?

2 points – **base excision repair.** "base repair" got 0.5 points.

What enzymes are involved?

4 points (one per enzyme) – **uracil N-glycosylase, AP endonuclease, DNA polymerase I** (the I is not required) **and DNA ligase.** Exonucleases may act between AP endonuclease and DNA polymerase, but you did not need to mention them and you did need to mention AP endonuclease, which recognizes the AP site from which the uracil base has been excised.

Would a similar mechanism work if the genome were RNA? Why or why not?

3 points – **No.** The uracil N-glycosylase removes any U, so no such activity is possible for a double-stranded RNA genome, which would have many U residues.

2. (10 points) Define microsatellite markers.

2 points – **Polymorphic sites containing a variable number of copies of a short (1-6 bp.) repeat. The allelic state of the marker is determined by the number of copies of the repeat.**

Explain how they are normally scored

4 points – **The site is amplified using unique flanking primers. The length of the PCR product is determined by the number of copies of the repeat.** The length of the PCR product is determined by running a gel, but "gel electrophoresis" is not an adequate answer (it was given 1 point).

For what purpose(s) they are normally used and their advantages and disadvantages (relative to SNPs) as genetic markers.

4 points – **They are used as genetic markers for mapping genes. Their main advantage is their high level of variability (heterozygosity), which allows multiple alleles to be tracked within a pedigree. Their disadvantages relative to SNPs are that they occur less frequently in the genome and they have a relatively high mutation rate (about 1 change per 1,000 generations).**

3. (10 points) Define LTR.

3 points – **Long Terminal Repeat.**

What sort of genetic element has LTRs?

3 points – Some retroelements (retroviruses and some transposable elements; the book uses the term "retroposon" but I don't). LINES do not have LTRs. DNA elements do not have LTRs. Inverted repeats and target site duplications are both different from LTRs.

What function(s) do they serve?

4 points – The LTR contains the start site of transcription and a polyadenylation site; these two sites determine the end of the "genomic" RNA. The LTR also contains sequences involved in integration, including short inverted repeats, but does not contain the primer-binding site, RNase H cleavage site, or genes for reverse transcriptase or integrase.

(3 points each for questions 4-18). In each of the following there are two or more statements. One is true (generally, it is taken directly from your textbook or another reliable source) and the others have been modified so as to be untrue, misleading, false, wrong, unlikely or bogus. Circle, check or otherwise designate the correct statement. Ambiguous marks (checking both, placing a mark between the two statements, etc.) will be considered wrong.

4. **b)** A typical bacterial genome is a single **circular** molecule.
5. **a)** In *E. coli*, *hisA* and *hisB* are two **genes**.
6. **b)** The bacterial origin of DNA replication can "fire" **before** the prior round of DNA replication is complete.
7. **b)** A single base mismatch in an allele-specific primer used in **PCR amplification** will provide maximal discrimination when located **at the 3' end** of the primer.
8. **a)** A single base mismatch in an allele-specific oligonucleotide used in **hybridization** will provide maximal discrimination when located **in the center** of the oligonucleotide.
9. **c)** A comparison of two haploid human genomes would reveal about **3 million** allelic differences.
10. **b)** New methods for the rapid and inexpensive sequencing of entire genomes are under active development and **dramatic improvements in technology have taken place in the past year and continue to take place**.
11. **a)** The **shotgun sequencing** protocol is a reasonable and commonly used approach to sequence whole genomes.
12. **b)** Only a **small** fraction of the human genome (about **5%**) is conserved with non-primate mammalian genomes (such as the dog).
13. **a)** Of that fraction of the human genome that is conserved with non-primate mammalian genomes (such as the dog), most (about 4/5) is **noncoding**.

14. **a)** A target site duplication is a duplication of sequence **present at the target site prior to the insertion** of a transposable element. This duplication is caused by insertion of the element and the two duplicates flank the inserted element.
15. **b)** The human alpha globin gene and the mouse beta globin gene are **paralogs**. They diverged when the alpha and beta globin genes became distinct, before the last common ancestor of humans and mice.
16. **a)** The human alpha globin gene and the mouse beta globin gene are **homologs**. Both parlogs and orthologs are homologs.
17. **a)** The basic motif DNA-binding domain in the Myc-Max system functions only in **dimeric** proteins.
18. **b)** RNA polymerase **II** transcribes protein-coding genes and miRNAs.

19. (6 points) You have a piece of DNA that includes the sequence:

5' GATGAGGATGAGGAGAAGTACCGGCCGCCCTGCGCATCACAATATGTTTCAGT 3'

To amplify this DNA by PCR you would use a pair of primers containing which **two** of the following eight primers (you must get both primers right to get any credit -- circle **two** letters):

- b) 5' - GATGAGGATGAGGAGAAG- 3'
e) 5' - ACTGAACATATTGTGATG- 3'

Each of the following statements applies to just one of the following three distinct types of cell: **a)** F⁺, **b)** F' or **c)** Hfr. Indicate which one by writing **a**, **b** or **c**. One cell type might be the answer for multiple questions. (2 points each)

20. These cells are used in interrupted mating experiments that can be used to map bacterial genes, generating a genetic map measured in minutes.

c) Hfr

21. These cells contain the F plasmid, an extrachromosomal circle of double-stranded DNA that is approximately 100 kb. in length.

a) F[±]

22. These cells contain the F plasmid integrated into a single site on the bacterial chromosome.

c) Hfr

23. These cells transfer a few bacterial genes very efficiently, generating stable partial diploids.

b) F'

24. (6 points) You isolate DNA from a plant and purify organellar DNA (DNA from the chloroplast and mitochondria, free of nuclear DNA). The mixture has equal amounts by mass of chloroplast DNA and mitochondrial DNA (two milligrams of each). You take a sample of the DNA, shear it by sonication and heat the mixture to denature it. You measure hybridization rate as the time required for half of the DNA to become double-stranded (not the initial reaction rate). You observe that the mitochondrial DNA hybridizes much faster; it takes only 1/4 as long to hybridize. Which is likely to be the case?

___d) The mitochondrial genome is 1/4 the size of the chloroplast genome.

The **rate** of hybridization ($1/t$) is proportional to the amount of complementary DNA present. Thus, there is more mitochondrial DNA of any given sequence present in the reaction. Given that you know that the mass of DNA is the same, the mitochondrial genome must be shorter (more copies per mass of DNA). This also follows from the relationship given in class, that **complexity is proportional to concentration (of nucleotides), and to the time**. Here, concentration in nucleotides is the same for chloroplast and mitochondrial DNA but the time is less for the mitochondrial DNA, meaning that the complexity must be (proportionally) less as well.