

Genomics Midterm Thursday, March 24, 2011

NAME: _____ STUDENT ID: _____

Take-home exam: Due Monday, March 29, 2011.

Return an edited version of this WORD document with answers by email to David Pollock (David.Pollock@ucdenver.edu) by **12:00 pm (noon) on March 29, 2011**. Please use “**Genomics Midterm**” as the subject.

I will send out a **confirmation** to everyone from whom I receive a completed exam. If you do not receive this confirmation, please call to find out why not.

All answers to questions should be typed and submitted as a WORD document. Include your name in the name of the document, *e.g.*, **DPollock_GenomicsMidterm.doc**

Answers to each question should be **concise yet informative and ~1/2 page**. For each question you answer, you should go over your notes and any reading related to the relevant lecture(s) thoroughly so that you fully understand the material and this understanding is reflected in your answer. The bonus question only requires a very short answer. Please do not re-order or re-number the questions.

If there are any necessary addendums, modifications, or further explanations that are required they will be posted on the class web site (at evolutionarygenomics.com).

Please **answer 9 out of 13** of the following questions (your choice):

1. Briefly explain how the presence of Segmental Duplications (SDs) can complicate mapping and assembly of short reads available from non-Sanger sequencing technologies. Describe a method by which you could ensure accurate mapping of SD paralogs.
2. Now that we know over 1000 complete human genome sequences, what is the utility of sequencing more genomes from different species, some closely related to humans, others distantly related. Specifically, provide four brief examples of how comparative genomics (and increased knowledge of genomic diversity) contributes to a greater understanding and/or annotation of the human genome. For each example, also identify which types of species (how closely related to humans, in a very rough sense) are most important for each comparison discussed. Each example, in total, should be no more than two-three sentences.

3. Why was “physical” mapping of genomes necessary for the clone-based DNA sequencing approach that has provided nearly all genome sequences? Will physical mapping of genomes remain necessary in this age of high-throughput whole-genome shotgun sequencing? Why or why not?

4. Based on Thomas et al. (2003)*, a) what conclusions can be drawn from Figure 1A + 1B; b) what conclusions are drawn from Figure 2A-2B; c) what conclusions arise from Figure 3A. Write 2-3 sentences for each question.
 *These figures were also covered in the Castoe lecture. Thomas et al., 2003, *Nature*, "Comparative analysis of multi-species sequences from targeted genomic regions".

5. List the advantages and disadvantages of array-based comparative genomic hybridization (array CGH) as a method of genome analysis for comparison within the human population as well as between human and other primate species.

6. What are the possible fates of novel alleles introduced into a population? Discuss the forces that determine allele fate for deleterious, neutral, and adaptive alleles.

7. When sequencing diploid individuals, each base must have 8x or greater coverage to have confidence in your base calling. If you sequence a genome at an average of 20x coverage, what fraction of the bases are at 8x or greater coverage? Use the Poisson distribution: $f(n; \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$, to calculate your answer.

8. How would you test whether a population is ideal (that is, whether it deviates from random mixing because it is, for example, highly inbred, or there are actually two source populations)? Please provide a full description in terms of observations, expectations, allele and genotype frequencies, and how they would be calculated.

9. Which type of transposable element would be best suited for building a novel regulatory network? Devise a new TE-derived regulatory network, outlining the steps needed for the process of exaptation of an active, mobile TE to the final regulatory network.

10. Describe briefly how conserved regions can be detected using comparative sequence data. How do the detection of conserved sites and conserved regions differ? What experimental design factors can be manipulated to improve the resolution of short regions?

11. Describe the major advances that enable “NextGen” DNA sequencing technologies to provide *much* higher sequence throughput than “Sanger”-based methods. Chemical reactions are never 100% efficient. How does this limit the popular Illumina sequencing technology? How does it limit the single-molecule DNA sequencing approaches?

12. Discuss reasons that we might want to understand polymorphism. Why is genetic drift a “force to be reckoned with”? Please couch your answer in terms relating the ill effects of inbreeding to “remote inbreeding” and expected amounts of variation in a population (disregarding selection).

13. Discuss the biological foundation of the “master element” model of transposable elements, and its importance for analyzing the evolutionary history of these elements.

Bonus Question: What is the sum of the genotype frequencies in a population?