

PLNT4610 BIOINFORMATICS

FINAL EXAMINATION

13:30 to 15:30 Monday December 8, 2014

Answer any combination of questions totalling to exactly 100 points. The questions on the exam sheet total to 120 points. If you answer questions totalling more than 100 points, answers will be discarded at random until the total points equal 100. This exam is worth 20% of the course grade.

Hand in the question sheets along with your exam booklet. All questions must be answered in the exam book. The question sheets will be shredded after the exam.

1. (10 points) When a multiple sequence alignment contains large gaps, what problem does this create for constructing phylogenetic trees? What is one solution to constructing phylogenies, when the alignment contains large gaps?

2. (10 points) List five sources of experimental variance in microarray experiments. Use point form for your answer.

3. (10 points) What is the basis for polymorphism with Microsatellite markers? Why are microsatellite markers usually more informative than markers such as RFLPs, RAPDS or SCARS, that only have 2 possible states?

4. (10 points) Fill in the blanks. Just write the answers for a - e on your answer sheet. Don't bother copying all paragraphs.

Camin-Sokal parsimony (0 → 1 only occurs)

Camin-Sokal further assumes that loss of the allele does not occur. It is therefore probably more appropriate for use with _____ a _____ rather than molecular marker data.

Dollo Parsimony (1 → 0 _____ b _____ 0 → 1)

Dollo parsimony assumes 0 as an ancestral state. It assumes **1 → 0** _____ b _____ **0 → 1**, but that both are rare over the evolutionary timescale being studied. Usually the most realistic method for population data.

Wagner Parsimony (1 → 0 = 0 → 1)

Wagner parsimony assumes that ancestral states are unknown, and that roughly equal rates of substitutions can occur in either direction. This assumption is probably not valid for most molecular marker methods, but would be appropriate for _____ c _____.

Polymorphism parsimony (1 → 0 > loss of an allele >> 0 → 1)

The above methods assume that polymorphism for a character is not retained in the population, but rather that one allele becomes fixed. In the absence of selection, the laws of population genetics predict that where two alleles exist in a population, one will be lost by random chance, especially when population size is small. This is what statisticians refer to as _____ d _____.

Felsenstein has proposed a model in which polymorphism can be retained in the population, thus effectively allowing what looks like _____ e _____ if the **1** allele becomes fixed at some later time.

8. (15 points) The following database classes or objects are examples of bad database practices. Explain what the problem is in each object or class, and what would be a better way to implement it.

a) The class is fine. What's wrong with the object, and how would you fix the problem?

<u>CLASS</u>		<u>OBJECT</u>	
Author		Schmidlup CT et al.	
Publication	?Publication	Publication	Population diversity in Ulm... Evidence for balancing sele...

b) A class was designed that would record the conditions for testing plants for disease resistance to microbial pathogens. In the example, two canola lines (Westar and Glacier) were inoculated with two different strains of the blackleg fungus (PG1, PG2), for a total of four experiments. The same concentration of inoculum was used in all experiments. Westar is resistant to PG1 but susceptible to PG2, while Glacier is resistant to both.

<u>CLASS</u>		<u>OBJECT</u>	
Experiment		GN285	
Pathogen	?Strain	Pathogen	PG1
Host	?Plant_line		PG2
Inoculum [sp/ml]	Float	Host	Westar
Disease score UNIQUE	Resistant		Glacier
	Susceptible	Inoculum [sp/ml]	10e6

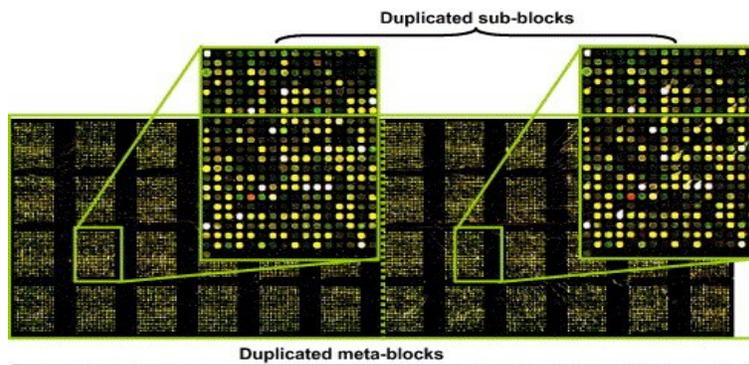
What is wrong with Experiment object GN285? Using the same Experiment class with no changes, how would you create objects that more accurately describe the four experiments?

9. (10 points) List 5 reasons why RNA-seq is more difficult with eukaryotes, as compared to prokaryotes

10. (5 points) For the following equation used in RNA-seq, define F, L and N.
What does FPKM stand for?

$$\text{FPKM} = F/LN$$

11. (5 points) Most microarrays today are designed so that redundant sets of genes are represented in at least two different regions of the array. An example is shown below. Explain the reason for this redundancy.



12. (5 points) In microarray and RNA-seq experiments, what is the advantage to doing larger numbers of biological replicates?