

PLNT4610 BIOINFORMATICS

FINAL EXAMINATION

1:00 p.m. to 3:00 p.m.

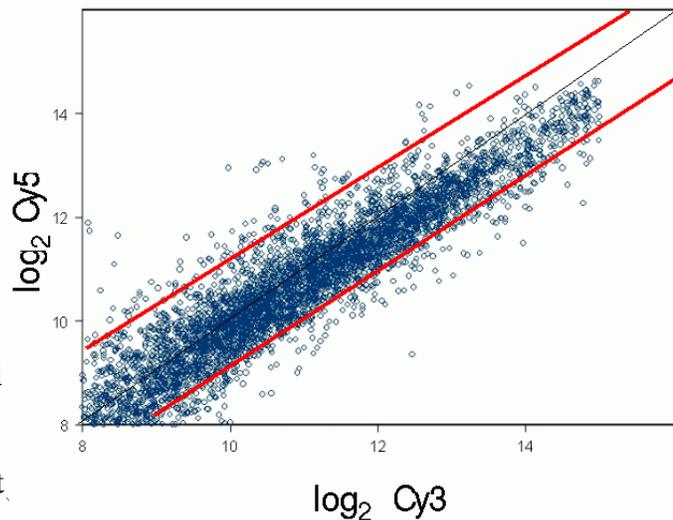
Tuesday December 22, 2009

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

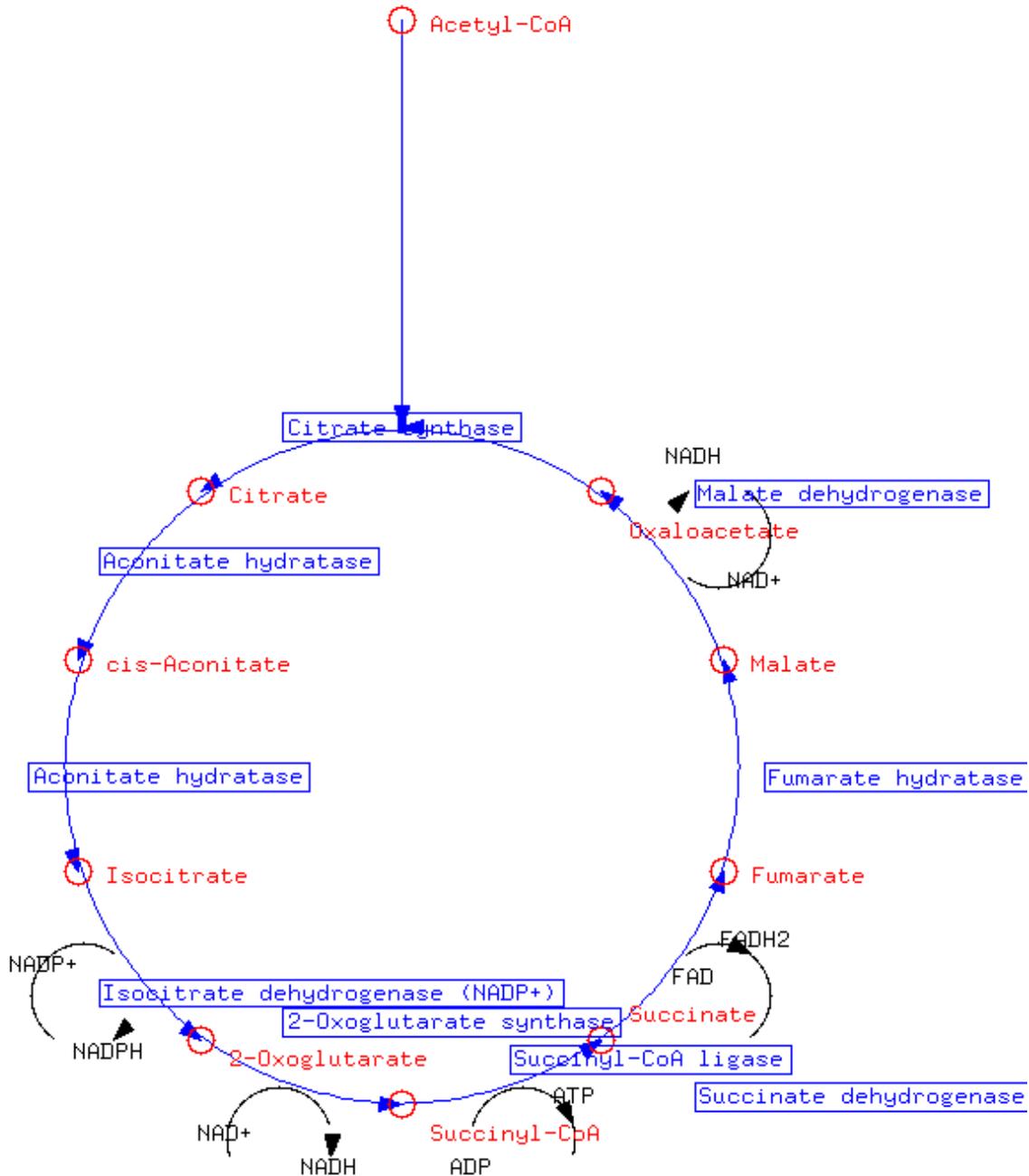
1. (10 points) Define the database term "class". What are the two main components of a class? How is a class distinct from an object? In this context, what are relations?

2. (10 points) In many microarray projects, an experiment is repeated, and the mRNA from one replicate is labeled with Cy5, and mRNA from a second replicate is labeled with Cy3. Both samples are mixed and allowed to hybridize with a single gene chip. The accompanying figure shows a plot for Cy5 signal plotted against Cy3 signal for a single treatment. Each point in the graph represents a single gene. Explain the purpose of such plots, and what they tell us.



Note that data points are more spread out at low Cy5/Cy3 values, and are closer together at high Cy5/Cy3 values. What is the reason for that difference?

3. (20 points) Draw a schema that could be used to implement a database of biochemical pathways. To help you visualize what you are trying to model, the TCA cycle is shown below as an example of a biochemical pathway. The object is to use a very small number of well designed classes that cleanly describe the components of a biochemical pathway and their relationships.



4. (20 points) Describe briefly how phylogenetic trees are constructed using distance methods and maximum parsimony. List at least one advantage and disadvantage of each.

5. (10 points) Data for molecular markers is typically represented using the symbols 1 to represent presence of a band, and 0 to represent absence of a band. When molecular marker data is used to construct phylogenies, the assumption is usually that a 1 --> 0 mutation is far more likely than a 0 --> 1 mutation. Explain the reasoning behind this assumption.

6. (10 points) Explain how the following equation is used in the Maximum Likelihood method for calculating the linkage distances θ from the observed frequencies of progeny, where

- f_i is the observed frequency of phenotype i
- N is the total number of progeny
- $P_i(R|G)$ is the probability that an individual of a certain phenotypic class is recombinant, given the observed phenotype

$$\theta^{new} = \frac{1}{N} \sum f_i P_i(R|G)$$

7. (10 points) What is the client/server model of computing? Give an example of a client program.

8. (10 points) Define web services and what they do. What are some of the advantages and disadvantages of using web services for data analysis, as opposed to programs running on one's local computer?

9. (10 points) Despite the fact that they organize data into information, web sites cannot be considered databases, in the formal sense. Explain why.

10. (10 points) Compare and contrast the use of molecular markers, DNA sequences, and protein sequences for constructing phylogenies:

- for different populations of the same species
- for different species that are closely-related
- for different species that are distantly-related