

**BioE/MCB/PMB C146/246, Spring 2005**

**Problem Set 5**

Due 21 February 2005 by midnight to [lfl@compbio.berkeley.edu](mailto:lfl@compbio.berkeley.edu). Remember to hand in problem set 4 as well!

1. 20 points

Why is it necessary to use masking for low complexity regions? Why is it necessary to use masking for coiled coil regions? Describe one other example of a feature that can confuse alignments in a similar way.

2. 20 points

Infer the function of the unknown genes below, given the phylogenetic profiles of their orthologs in various species. Explain your decision; do not just state the function. How confident are you of these predictions?

<i>Species Abbreviations</i>							<i>Name</i>	<i>Function</i>
<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>J</i>	<i>R</i>	<i>S</i>		
1	1	1	0	1	1	1	CheZ	Chemotaxis
1	1	1	0	1	1	1	CheY	Chemotaxis
1	1	1	1	1	0	1	EnvZ	Histidine kinase
1	0	1	1	0	0	0	FimA	Type I pilin
1	0	1	1	0	0	0	FimG	Type I pilin
1	0	1	0	1	0	0	PilA	Type IV pilin
1	1	1	0	1	1	1	Unknown 1	
1	0	1	1	1	0	0	Unknown 2	

3. 20 points

Infer the functions of the unknown protein-coding genes, using a domain fusion approach. Explain your decision; do not just state the function.

<i>Domain Abbreviations</i>					<i>Species</i>	<i>Name</i>	<i>Function</i>
<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>			
0	0	1	0	0	<i>A. nidulans</i>	CPSase	Carbamoyl-phosphate synthetase
0	0	0	1	0	<i>A. nidulans</i>	ATCase	Aspartate transcarbamylase
0	0	0	0	1	<i>A. nidulans</i>	DHOase	Dihydroorotase
1	0	0	0	0	<i>E. coli</i>	TrpC	Tryptophan biosynthesis
0	1	0	0	0	<i>E. coli</i>	TrpG	Tryptophan biosynthesis
1	1	0	0	0	<i>S. cerevisiae</i>	Unknown 1	
1	0	1	1	1	<i>D. melanogaster</i>	Unknown 2	

4. 20 points

- (A) What are sources of errors in functional annotation of protein sequences? How has the situation improved and worsened since Prof. Brenner's 1998 paper?
- (B) A BLASTP search was conducted on a hypothetical open reading frame. Given the BLAST results and Gene Ontology assignments for the top hits on the next page, what can you deduce about the functional roles of the unannotated query protein?

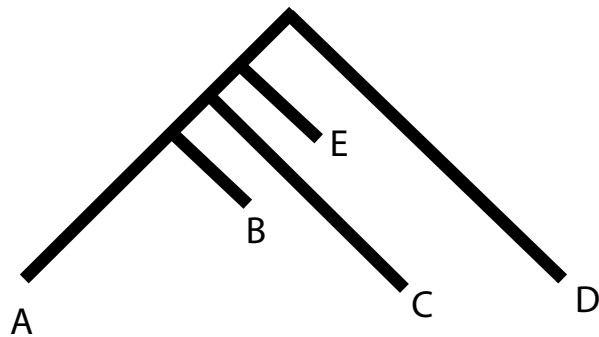
5. 20 points

You would like to know the function of an unknown protein, E, found in *P. falciparum*, the malaria parasite (a single-celled eukaryotic "protist"). You know the protein has significant BLAST matches to proteins A, B, C, and D.

Query sequence: E

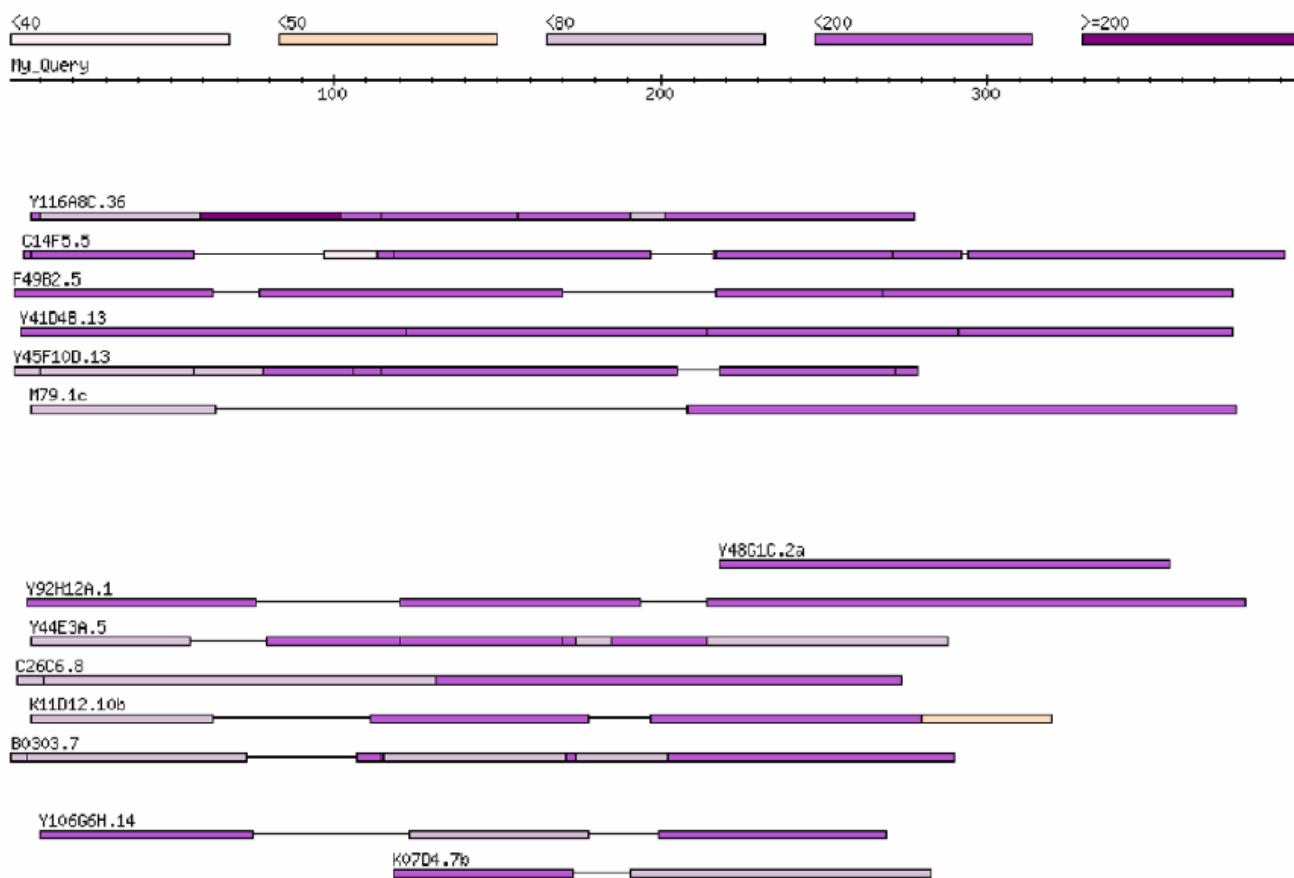
Protein	E-value	Function	Species
B	1e-45	green	<i>S. cerevisiae</i>
A	1e-41	green	<i>C. elegans</i>
C	1e-38	purple	<i>C. elegans</i>
D	1e-22	purple	<i>E. coli</i>

The sequences are related by the following tree:



- (A) What is the best function assignment for protein E?
- (B) What pair(s) among A, B, C, D, and E are most likely to be paralogs? Where did a duplication most likely occur in the tree? What species might have lost gene(s)?

## WU-BLAST RESULTS



High Score	P <sub>N</sub>	# HSP	Biological Process	Cellular Component	Molecular Function	Evidence
229	4.60E-17	9	Hydrogen transport	unclassified	ATP binding, Ca <sup>2+</sup> binding	Inferred by electronic annotation
140	9.30E-17	8	Intracellular signaling cascade	unclassified	unclassified	Inferred by electronic annotation
172	3.30E-11	4	Intracellular signaling cascade	unclassified	ATP binding, protein tyrosine kinase	Inferred by electronic annotation
162	9.00E-11	4	Intracellular signaling cascade, programmed cell death	plasma membrane	SH3/SH2 adaptor protein	Inferred by mutant phenotype & sequence similarity
126	2.70E-10	8	Unclassified	unclassified	unclassified	
164	9.10E-10	2	Intracellular signaling cascade	unclassified	ATP binding, protein tyrosine kinase	Inferred by electronic annotation
160	1.20E-09	1	Intracellular signaling cascade	unclassified	ATP binding, protein tyrosine kinase	Inferred by electronic annotation
154	3.80E-09	3	Unclassified	unclassified	unclassified	
121	4.60E-09	6	Unclassified	unclassified	unclassified	
122	1.80E-06	3	Unclassified	unclassified	unclassified	
130	2.20E-06	4	protein amino acid phosphorylation	unclassified	ATP binding, protein serine/threonine kinase	Inferred by electronic annotation
96	2.40E-06	7	Unclassified	unclassified	unclassified	
119	4.20E-06	3	Unclassified	unclassified	unclassified	
124	1.00E-05	2	Unclassified	unclassified	unclassified	