

Genomes and Genomics 2010 Practical Assessed Writeup 2010

In your writeup, which should not need to be longer than 8 printed pages, you should

1 Describe the process by which you isolated the EST sequences.

Note that this should include comments on why things might have failed (if you had failures).

This is expected to cover about 1 page. 15% of marks

2 Draw up "molecular CVs" giving the information below (a-k) for each of your sequences.

You may not have information for some sequences (e.g. if the reaction failed) and you should therefore enter what information you have. If you used sequences from clones other than the ones you processed, you may not have some data (such as insert length estimated from the gel) but please provide what data you can from the sequencing chromatogram. Remember to include quality data supporting each of your annotation statements.

a Clone name

b Insert length

c Sequence length acquired and comments on sequence quality

d Position of start of putative open reading frame or initiation methionine; length of open reading frame

e Position of polyA tail (if present)

f Best BLAST matches in nonredundant database

g Best BLAST match in annelid sequences

h Best BLAST match in nematode sequences

i Match to InterPro domains

j A description of what the function of the gene is likely to be in *Euperipatoides kanagrensis*. In this section you should state the pieces of evidence that lead you to make this statement of putative function. Please give literature references where relevant.

k AND, FINALLY in less than 30 words, provide a summary of this putative functional annotation that could be included in the 'definition line' for the sequence in GenBank (we ask for less than 30 words because this is how genes are annotated on GenBank - not by essays, but by short, summary statements)

This is expected to cover about 4 pages. 35% of marks

3 Discuss whether your sequences offer support for the ARTICULATA or ECDYSOZOA hypotheses of onychophoran relationships.

a Describe how you can use your sequences to answer this difficult evolutionary question. (Why is this an interesting question to answer? Why might sequences be able to give a clearer answer than can morphological comparisons?)

b When you compare your sequence to those available for Nematodes or Annelids, which is it most similar to, and what model of evolutionary relationships does this finding support?

This is expected to cover about 1 page. 20% of marks

4 Answer the following questions (a-d).

a Why is an EST strategy useful for "neglected" organisms such as *E. kanagrensis* ?

b Why is an EST strategy **also** useful for model organisms such as humans where the whole genome has been sequenced?

c Based on what we know about other animals, we predict that *E. kanagrensis* will have ~18000 protein coding genes (with a mean length of 1.1 kb per mRNA). If we want to identify as many *E. kanagrensis* genes as we can, how many ESTs should we aim to sequence from how many cDNA libraries?

d The *E. kanagrensis* genome is estimated to be ~4.7 Gigabases in size. Would sequencing the whole *E. kanagrensis* genome be a more efficient strategy to identify all the protein coding genes?

This section is expected to cover about 2 pages. 30% of marks

We expect you to give references to relevant literature, and also to any information you have derived from expert web sites.

The writeup should be handed in to the relevant BOX by 4 pm on Thursday 18th March 2010.

Please "name" your writeup with your matriculation number only.