

The Power of Computers in Biology

Worksheet

Task A: Identify the mysterious “Nucleotide Sequence R”

Question 1

Sequence R has an excellent match to a known protein, indicated by its low E-value and high percentage identity. We assume **Sequence R** codes for this protein.

What is the name of this protein?

Hint: To find out more about this protein, click the link under “Accession”. Then you will see the protein name near the top of the page, in bold.

Question 2

In which organism is this protein found?

Hint: Look for the line beginning “SOURCE”.

Question 3

What is the biological role of the protein?

Hint: Do a Web search for the name of the protein (from your answer to Question 1).

Task B: Search for a match to Sequence R in the human genome

Question 4

On which human chromosome has the best match been found?

Question 5

In the alignment between **Sequence R** and the human genomic DNA, can you see evidence of a substitution mutation?

If so: sketch the region that includes the substitution.

Question 6

Can you see evidence of an insertion or deletion mutation?

If so: sketch the region that includes the insertion or deletion.

Question 7

Do you think the human genome includes a functional version of **Sequence R**?

Explain your answer.

Question 8

Sequence R comes from the house mouse and codes for L-gulonolactone oxidase, an enzyme that synthesizes vitamin C.

Vitamin C is vital for both humans and mice.

Does your answer to Question 7 tell us anything about how the diet of humans might differ from the diet of mice?

Task C: Using BLAST at the command line

Question 9

Which animals did you analyse?

Do you think these animals can produce their own Vitamin C?

Did any of the results surprise you?

Daniel Barker, Heleen Plaisier, Laura CE Campbell, Stevie A Bain, Richard Fitzpatrick and Chenxi Zhang

[4273pi Bioinformatics Education Project](#)

Copyright and related rights waived via [CC0 1.0 Public Domain Dedication](#).

Version 3.0