Bioinformatics: Food Detective

Worksheet

TASK ONE

|  |
| --- |
| * Open a Web browser and type the following address into the address bar: [**4273pi.org**](https://4273pi.org) |
|  |
|  |
| * Click on the **‘Students’** tab. Under **‘Bioinformatics – Food Detective’**, view sequences for the **extended** workshop. These contain A, C, T and G for the four bases of DNA and a one-line heading, like this: |
|  |
| * Keep this open throughout the workshop so you can copy and paste the sequences. |

To identify the animal species to which these sequences belong, you will search for sequences in the NCBI database which show high similarity to each DNA barcode sequence.

The NCBI maintains a huge database containing almost all known DNA sequences and is free to use.

|  |
| --- |
| 1. Open a new tab in your Web browser and search for **NCBI BLAST**. *This is the DNA search engine we will use.* 2. Click the link **BLAST: Basic Local Alignment Search Tool**. 3. Click **Nucleotide BLAST**. 4. Paste Sequence A into the **Enter Query sequence** box at the top of the page. 5. Scroll down to **Optimize for** and choose **somewhat similar sequences**. *This causes BLAST to do a slow and careful search.* 6. Scroll to the bottom of the page and click **BLAST** |
|  |

BLAST may take a few minutes to run, during which time ‘Status’ on the Web page is ‘Searching’. Do not refresh the page! In a few moments, the results will appear on a webpage headed **‘BLAST > blastn suite > results for …’**.

|  |
| --- |
| 1. Wait for results to appear. 2. Scroll down a little to see which DNA sequences are similar to Sequence A. 3. The best-matching sequence is listed first. Write its scientific name in your Results table.     **Now, repeat steps 1-6 for sequences B-M.** |
|  |

## **Table 1:** Complete table using BLAST results

|  |  |  |  |
| --- | --- | --- | --- |
| **Sequence** | **Species – scientific name** | **Species – common name** | **E-value** |
| A |  |  |  |
| B |  |  |  |
| C |  |  |  |
| D |  |  |  |
| E |  |  |  |
| F |  |  |  |
| G |  |  |  |
| H |  |  |  |
| I |  |  |  |
| J |  |  |  |
| K |  |  |  |
| L |  |  |  |
| M |  |  |  |

## **Table 2:** Examples of scientific names

|  |  |
| --- | --- |
| **Scientific name** | **Common name** |
| *Bos taurus* | Cattle |
| *Gallus gallus* | Chicken |
| Makaira sp. | Marlin |
| *Homo sapiens* | Human |
| *Balaenoptera musculus* | Blue whale |
| *Epinephelus guttatus* | Red hind |
| *Ovis aries* | Sheep |
| *Scinax* sp. | Snouted treefrog |
| *Sus scrofa* | Pig |
| *Streptomyces* sp. | Bacteria |

|  |
| --- |
| **Question 1:** What do your results in Table 1 tell us about the DNA in the sausage? Does the meat seem to be 100% pork? |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

|  |
| --- |
| **Question 2:** Do any of your results seem unexpected? Explain. |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

|  |
| --- |
| **Question 3:** Are your results ***really*** unexpected? Think about how sausages are made and how DNA is extracted. |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

|  |
| --- |
| **Question 4:** Compare the E-values of sequences in Table 1 to each other. Are all of the results equally reliable? Are these results surprising? |
|  |
|  |
|  |
|  |
|  |
|  |

|  |
| --- |
| Stevie A Bain, Daniel Barker, Laura CE Campbell and Richard Fitzpatrick |
| 4273pi Bioinformatics Education Project, <https://4273pi.org> |
|  |
| Copyright and related rights waived via CC0 1.0 Public Domain Dedication |
| (<https://creativecommons.org/publicdomain/zero/1.0>). |
|  |
| **Version 2.2 (Extended)** |