Bioinformatics:

Food Detective

Worksheet

|  |
| --- |
| Stevie A Bain, Daniel Barker 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 2.1 (long)** |

# Introduction

In this workshop, you will be provided with DNA barcode sequences from a sausage, described as a ‘100% pork sausage’, bought from a butcher in the UK.

Your task is to see what meat the sausage **actually** contains. Is it really 100% pork (pig)? Or does it contain material from other animals as well?

You will compare DNA barcode sequences from the sausage[[1]](#footnote-1) to known DNA in the publicly available database. This will give us information about which animals’ DNA is in the sausage.

A DNA barcode is a region of DNA sequence that is common to all animals but varies between species. Think about barcodes in supermarkets: each item has a barcode but differences in these barcodes allow different items to be distinguished when scanned at the tills. QR codes also function in a similar way.

You will use the publicly available DNA sequence database at the National Center for Biotechnology Information (NCBI) to search for matches to these DNA barcodes. You will then be able to identify what animal DNA is present in the sausage.

Work through the task, writing your answers in this 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 and you will find a link to the sausage DNA barcode sequences. These contain A, C, T and G for the four bases of DNA and a one-line heading, like this: |
|  |
| >Sequence\_C GACGAGAAGACCCTATGGAGCTTTAACTAACCAACCCAAAGAGAATAGATTTAACCATTAAGGAATAACAACAATCTCCATGAGTTGGTAGTTTCGGTTGGGGCGACCTCGGA   * Keep this open throughout the workshop so you can copy and paste the sequences when required. |

**For Task One, you will be using Sequences A to H.**

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. It can be searched using software called BLAST.

|  |
| --- |
| 1. Open a new tab in your Web browser and go to: [**https://blast.ncbi.nlm.nih.gov**](https://blast.ncbi.nlm.nih.gov) |
| 1. Click ‘Nucleotide BLAST’. |
|  |
| 1. Paste ‘Sequence\_A’ into the box at the top of the page labelled ‘Enter Query sequence’. It is helpful to paste the one-line heading as well as the sequence. |
|  |
| 1. Select ‘somewhat similar sequences’ under ‘Optimize for’ |
|  |
| 1. 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. The first result in the table is the best match to Sequence\_A in the database. Make a note of the species this best match comes from, in Table 1 on the next page. Be sure to also record the E-value. |
| 1. Repeat the BLAST search for Sequences B to H and complete Table 1. |

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Sequence** | **Species – *scientific name*** | **Species – common name** | **E-value** |
| A |  |  |  |
| B |  |  |  |
| C |  |  |  |
| D |  |  |  |
| E |  |  |  |
| F |  |  |  |
| G |  |  |  |
| H |  |  |  |

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

|  |  |
| --- | --- |
| **Scientific name** | **Common name** |
| *Bos taurus* | Cattle |
| *Gallus gallus* | Chicken |
| *Hollandichthys multifasciatus* | Characin |
| *Homo sapiens* | Human |
| *Balaenoptera musculus* | Blue whale |
| *Epinephelus guttatus* | Red hind |
| *Ovis aries* | Sheep |
| *Scinax* sp. | Snouted treefrog |
| *Sus scrofa* | Pig |
| *Viola odorata* | Sweet violet |

|  |
| --- |
| **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. |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

TASK TWO

Follow **Steps 1-10** again but this time search for sequences I to M.

## **Table 3:** Complete table using BLAST results for sequences I - M

|  |  |  |  |
| --- | --- | --- | --- |
| **Sequence** | **Species – scientific name** | **Species – common name** | **E-value** |
| I |  |  |  |
| J |  |  |  |
| K |  |  |  |
| L |  |  |  |
| M |  |  |  |

|  |
| --- |
| **Question 4:** Compare the E-values of sequences in Table 1 to those of sequences in Table 3. Which table contains sequences with the highest E‑values? |
|  |
|  |
|  |
|  |
|  |
|  |

|  |
| --- |
| **Question 5:** Which table allows us to identify the species the DNA sequences come from most reliably? Why? |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

1. Most sequences are **real** sequences from the sausage. In one case we have substituted a publicly available sequence instead. [↑](#footnote-ref-1)