Bioinformatics: Food Detective

*Answers*

|  |
| --- |
| 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>). |
|  |
| **For Version 2.0 (long)** |

# TASK ONE

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

*Note: we have reformatted E-values. ‘1 x 10–48’ will appear as ‘1e-48’ in BLAST output. For our purposes, either format is fine.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Sequence** | **Species – *scientific name*** | **Species – common name** | **E-value** |
| A | ***Sus******scrofa*** | **Pig** | **2 x 10–48** |
| B | ***Sus******scrofa*** | **Pig** | **2 x 10–48** |
| C | ***Bos taurus*** | **Cattle** | **9 x 10–47** |
| D | ***Sus******scrofa*** | **Pig** | **5 x 10–50** |
| E | ***Ovis aries*** | **Sheep** | **9 x 10–47** |
| F | ***Sus scrofa*** | **Pig** | **5 x 10–50** |
| G | ***Homo sapiens*** | **Human** | **8 x 10–48** |
| H | ***Gallus gallus*** | **Chicken** | **2 x 10–48** |

Note: E-values may vary slightly, due to daily updates to the sequence database.

**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?

**The meat is not absolutely 100% pork (pig). At least some DNA of cattle, sheep and human is also present in the sausage.**

**QUESTION 2:** Do any of your results seem unexpected? Explain.

**Yes, since the sausage is 100% pork, we expected only pig DNA to be present. Instead, we found DNA sequences for sheep, cattle, chicken and human as well.**

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

**It is not too surprising that other species’ DNA sequences are found in the sausage. Chicken, cattle and sheep products (i.e. chicken, beef and lamb) are also prepared and stored in the butcher’s shop. Also, sausage meat is commonly mixed by hand, potentially explaining the traces of human DNA. Humans extracted the DNA from the sausage, so there may have been some contamination with human DNA during that process. Even the smallest traces of DNA can be picked up in the DNA extraction and sequencing processes.**

TASK TWO

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Sequence** | **Species – *scientific name*** | **Species – common name** | **E-value** |
| I | ***Neomerinthe hemingwayi*** | **Spinycheek scorpionfish** | **0.004** |
| J | ***Balaenoptera musculus*** | **Blue whale** | **0.027** |
| K | ***Streptomyces* sp.** | **Bacterium** | **0.009** |
| L | **No match** | **No match** | **No match** |
| M | ***Scomber scombrus*** | **Atlantic mackerel** | **0.002** |

Note: Species and E-values may vary, due to daily updates to the sequence database.

**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?

**The E-values in Table 3 are higher than those in Table 1.**

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

**Table 1 has more reliable BLAST results than Table 3. This is because the E-values in Table 1 are lower than those in Table 3.**

**The E-value is how many times we would expect to see a match of this quality, between our sequence and a sequence in the database, by chance alone.**

* **If the E-value is high, the match is unreliable.**
* **If the E-value is low, the match is reliable.**
* **If the E-value is 0, the match is extremely reliable.**