Worksheet 1 – Molecular Tools for Yeti Hunting

**Question 1.** In Sequence alignment A (above), part of the human 12S rDNA sequence is used as the Query and has been used to search Genbank - one alignment result is shown. What can you infer from these sequences and your knowledge of the 12S rDNA locus?

The sequences align at all of the bases. This should be interpreted as the sequences are a match and most likely the same species.

**Question 2.** In Sequence alignment B, the same two 12S rDNA sequences were aligned, but what is shown is a different region of the same gene. The differences are highlighted in the second (sbjct) sequence. What might you infer from this alignment?

This alignment suggests that the species are different having 5 mismatches/differences in a similar size fragment.

**Question 3.** While DNA isolation and a BLAST search is a powerful manner in which to identify a specimen, there can be accurate sequences that do not match anything in the database.

What are two possible explanations that you could use to explain the lack of a perfect match between your query sequence and the database?

One answer for the lack of a perfect match would be that there are errors in the sequence determined from the sample (query) or even possibly the database (this does happen). Ideally, this should be minimized through sequencing the sample multiple times and in both directions.

A second possibility would be that there is currently no match in the database, and that the query (sample) sequence represents either a mutated version of the sequence or identifies an unknown sample (not previously before been sequenced).

**Critical thinking question.** An alignment between the human 12S rDNA gene sequence (used in questions 1 and 2) with another Genbank file (EU626452) (alignment not shown) also produces a 100% match in the region shown in alignment A. However, EU626452 is identified in the database as a uncultured bacterial clone from a wetlands ecosystem sequencing project. Why do you think that this Genbank sequence identification may be incorrect?

The primary reason should be that prokaryotes do not have a mitochondrial genome (nor a mitochondrion) and therefore should not possess a 12S rRNA sequence.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Worksheet 2

**Question 1.**  Using the BLAST search and NCBI information, prepare a table of the results for Dr. Shipton.

|  |  |  |  |
| --- | --- | --- | --- |
| Sample | Accession | Description (record just genus and species here) | Identities (record as #/#) |
| Yeti1 | KJ155697.1 | *Ursus maritumus*  | 104/104 |
| Yeti2 | KJ155697.1 | *Ursus maritumus* | 104/104 |
| Yeti3 | MN181405 | *Canis lupus familaris* | 103/103 |
| Yeti4 | NC\_045205.1 | *Capricornis rubidus* | 104/104 |
| Yeti5 | MT048504.1 | *Bos taurus* | 104/104 |
| Yeti6 | AC025627.13 | *Homo sapiens* | 104/104 |

**Question 2**: Are any of the samples of unknown description (origin)?

All of the samples match (perfectly) at least one databank sample.

**Question 3**: Do any of the samples match the same genus/species description?

Yeti 1 and Yeti2 match the same GenBank sequence and are most likely duplicates.

**Question 4**: What is the most likely conclusion that you can make from the artifacts provided by Mr. Norgay?

Students should be able to conclude that the samples are of known origin.

**Question 5:** What are the common names of the origin of the samples? Hint: Use Google.com if you don’t know the common name.

|  |  |  |
| --- | --- | --- |
| **Sample** | **Genus/species (copy and paste from table above)** | **Common name** |
| Yeti1 | *Ursus maritumus*  | Polar bear |
| Yeti2 | *Ursus maritumus* | Polar bear |
| Yeti3 | *Canis lupus familaris* | Dog |
| Yeti4 | *Capricornis rubidus* | Red serow |
| Yeti5 | *Bos taurus* | Cow  |
| Yeti6 | *Homo sapiens* | Human |

The table below has been partially filled out since Yeti1 will be a 100% match to Yeti1 (and so on).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Yeti1 | Yeti2 | Yeti3 | Yeti4 | Yeti5 | Yeti6 |
| Yeti1 | 100% | 100% | 90% | 91% | 90% | 85% |
| Yeti2 |  | 100% | 90% | 91% | 90% | 85% |
| Yeti3 |  |  | 100% | 88% | 91% | n/a |
| Yeti4 |  |  |  | 100% | 90% | n/a |
| Yeti5 |  |  |  |  | 100% | 90% |
| Yeti6 |  |  |  |  |  | 100% |

**Question 6:** In some cases, there may not be a match provided. Why do you think this might be?

Sometimes the match is of lower quality (less perfect) and therefore BLAST does not create the alignment. You can change the settings prior to performing the search and receive matches. If you run the BLAST allowing more dissimilar matches, the sequences alignments will be presented.

**Question 7:** Based on the results of the table above and your knowledge of the animal origin of the samples, explain why it makes sense (or not) the percent similarity of the results. Is there anything unexpected in the results?

Since all of the samples were identified as mammals, the similarity is not surprising. The only unexpected results are that the Yeti6 sequence does not come up as a match for Yeti3 or Yeti4. Since Yeti6 was identified as human and Yeti3 is dog and Yeti4 is red serow, these are not unexpected results.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Worksheet 3**

**Question 1:** Look at the multiple sequence alignment in Figure 1. What can you infer from the results of the human (NC\_012920) sequence aligned with the other sequences?

If the students look at the sequence alignment, they should conclude that the human sequence is different at a number of bases as compared with the other sequences.

**Question 2:** If you look closely at the six non-human sequences, you should be able to determine that these sequences (99 nucleotides) are 100% identical. What might you be able to state about these results?

The information suggests that the five sequences represent the same species or at least a very closely related species.

**Question 3** Using the Genbank identification, fill out the table below to indicate the animal species that are represented in Figure 1.

|  |  |  |
| --- | --- | --- |
| **Genbank accession number** | **Genus and species** | **Common name** |
| MF579937 | *Mammuthus primigenius* | Woolly mammoth |
| KY616981 | *Loxodonta cyclotis* | African forest elephant |
| KY616977 | *Loxodonta africana* | African bush elephant |
| KY364233 | *Mammut americanum* | Mastodon |
| KY499558 | *Elephas antiquus* | Straight-tusked elephant |
| AJ428946 | *Elephas maximus* | Asian elephant |

**Question 4:** Based on your identifications, are the non-human animal 12S rDNA sequences aligned in Figure 1 representing identical species? If not, are they related? Are any of the species extinct?

These are all related (elephantids) species, but they are not identical. There are three living members in this group - the two African elephant species and the Asian elephant. There are three extinct species, including the woolly mammoth and mastodon.

**Question 5**: When the alignments are completed, copy and paste the MSA into the box below (Question 5). If the formatting does not look aligned, change the text to Courier font (a uniform font width) and set the size of the font to 9 or 10.

**Question 5:**

Yeti3 CTTAGCCCTAAACATAG-ATAATTTTACAACAAAATAATTCGCCAGAGGACTACTAGCAA

Yeti4 CTTAGCCCTAAACATAAATAATTGTAAAAACAAAATTATTCGCCAGAGTACTACCGGCAA

Yeti1 CTTAGCCTTAAACATAAATAATTTATTAAACAAAATTATTCGCCAGAGAACTACTAGCAA

Yeti2 CTTAGCCTTAAACATAAATAATTTATTAAACAAAATTATTCGCCAGAGAACTACTAGCAA

Yeti5 CTTAGCCCTAAACACAGATAATTACATAAACAAAATTATTCGCCAGAGTACTACTAGCAA

Yeti6 CTTAGCCCTAAACTCTAATAGTTACATTAACAAAACCATTCGCCAGAGTACTACAAGCAA

 \*\*\*\*\*\*\* \*\*\*\*\* \* \*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*\*\* \*\*\*\*\* \*\*\*\*

Yeti3 TAGCTTAAAACTCAAAGGACTTGGCGGTGCTTTATATCCCTCTA

Yeti4 CAGCCCAAAACTCAAAGGACTTGGCGGTGCTTTATATCCATCTA

Yeti1 CAGCTTAAAACTCAAAGGACTTGGCGGTGCTTTAAACCCTCCTA

Yeti2 CAGCTTAAAACTCAAAGGACTTGGCGGTGCTTTAAACCCTCCTA

Yeti5 CAGCTTAAAACTCAAAGGACTTGGCGGTGCTTTATATCCTTCTA

Yeti6 CAGCTTAAAACTCAAAGGACTTGGCAGTGCTTTATATCCCTCTA

 \*\*\* \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\*\* \* \*\* \*\*\*

**Question 6:** With the information that you have now generated in the MSA, compare the alignments. How similar/different are the sequences (keep in mind your % similarity results from the previous work)?

These sequences are not identical (except for Yeti1 and 2).

**Question 7:** Are there regions in this MSA that might be good indicators of where a variable region might be to distinguish sample identity? (Keep in mind that sometimes gaps shift the alignment and make a region look more variable than it is. Look at the sequence, not just the \*.)

There are differences throughout this entire region that could be important

**Question 8**: What can you tell Dr. Shipton when she calls about the data that you have so far collected?

Everything is identifiable as a specific species.

**Question 9**: Since Mr. Norgay provided you with 30 total specimens, how might you explain to Dr. Shipton about what you might be looking for in the remaining samples and those that she is sending to you?

Ideally, one or more specimen sequences will fail to be a 100% match in the 12S rRNA region that you are sequencing, suggesting a potential non-identified species has been sampled.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Worksheet 4**

**Question 1**: Can you identify the Kongma artifacts through sequencing of the 12S rRNA gene? If so, what might they be and why? *Remember in your analysis that synthetic DNA sequences are not representative of existing species.*

|  |  |
| --- | --- |
| Kongma1 | 96% (100 of 104) match to human; not a known species |
| Kongma2 | 100% (105 of 105) match to *Ailurus fulgens* (lesser (red) panda) - known species |
| Kongma3 | 100% (102 of 102) match to *Mutiacus vaginalis* (Indian mutjac), known species |

**Question 2**: What might you be able to talk with Dr. Shipton about regarding your results? Do you think that there might be other molecular steps that you could take to further your understanding of these results or to support any potential findings?

First, the students should suggest that the results need to be replicated.

Second, they might (should) suggest that other DNA barcodes be tested or to expand the 12S rRNA sequence that was analyzed here. Expanding what DNA barcodes are analyzed could provide more insight into the differences in the samples that might be more conclusive.