# Lab

| | |
|---|---|
| **Topic:** | Open Reference Frames (ORFs) |
| **Techniques:** | Use of NCBI resources |
| **Collaboration Policy:** | The lab should be completed **working in pairs** |
| **Submission Deadline:** | |

# Overview

**To receive full credit for this lab, you must get through all questions other than those labeled after the "Extra time?" prompt.**

---

Because it will be easier to properly predict genes for a prokaryote than a eukaryote, we are going to start with a prokaryote, such as *E. coli*. But since the full genome is rather large, we will also start with a smaller self-replicating DNA molecule known as a plasmid, specifically plasmid `pACYC184`.

1. Go to the NCBI database and search for pACYC184, and specifically the one described as "Cloning vector pACYC184".

   **Question:** How many nucleotides does this sequence have?

   **Question:** What is its accession number?

   **Question:** What are the first 10 nucleotides reported in its representation? (To be fair, this is actually a circular molecule, so the "start" is only by convention.)

2. Next, go to NCBI's ORF finder and enter the accession number for pACYC184, and have it compute all ORFs having minimum length of 150bp.

   **Question:** How many such ORFs are found?

   **Question:** How many nucleotides are in the longest ORF?

   **Question:** At what nucleotides does the longest ORF start and stop?

3. Click on the longest ORF to examine its details. Notice a box to the left that by default shows its amino acid sequence.

   **Question:** What are the first four amino acid characters?

4. You can switch to see the underlying nucelotide sequence by clicking on the "Display ORF as..." label.

   **Question:** What are the first 12 nucleotides?

   **Question:** Which of the stop codons ends this ORF?

5. Not every ORF is necessarily a gene. One way to suggest that an ORF is a gene is by comparing its sequence to a database of known genes from other genomes to look for similarity. BLAST is a popular such tool (and we will soon explore the underlying algorithm it uses for sequence alignment). The NCBI ORF Finder conveniently offers a button to perform a BLAST search for a selected ORF. (In fact, there is a "BLAST" button and a "SmartBLAST" button.) Let's use the SmartBLAST button.

   **Question:** What conclusion is suggested by a SmartBLAST on this ORF?

6. Let's go an examine the *second longest* of the identified ORFs.

   **Question:** How many nucleotides are in this ORF?

   **Question:** At what nucleotides does this ORF start and stop?

   **Question:** How would you interpret the fact that its start index is larger than its stop index?

   **Question:** How would you interpret the fact that its start index is larger than its stop index?

   **Question:** What conclusion is suggested by a SmartBLAST on this ORF?

7. Go back to the original database in which we found this plasmid. Within that view, there is a section labeled "FEATURES". Notice that two of those miscellaneous features are described as genes.

   **Question:** give the start..end indices and descriptions for the two genes.

   **Question:** Which of these corresponds to the longest ORF that we examined earlier?

   **Question:** Can you find an ORF that corresponds to the other of these identified genes?

   **Question:** What if we remind you that this was actually a circular molecule? Can you find a pair of ORFs that are reported by the ORF finder that together form this gene?

---

## Extra time?

Begin a similar such analysis on the guinea pig mDNA, to see whether the longest ORFs correspond with identified genes.

---

Last modified: Sunday, 11 February 2018

---