

Bioinformatics Exercise

1. extract_sequences Program

The `extract_sequences` program is available on Ron for you to use. It will take a pattern (as the first argument) and a FASTA-formatted file (as the second argument) and print all sequences whose headers match the pattern. For instance, to return all amino acid sequences with "oxidase" in the name, you can do the following:

```
extract_sequences "oxidase" myproteins.faa
```

Where `myproteins.faa` is some FASTA file that contains amino acid sequences.

If you would like to save the output of this program, you can redirect it to a filename. Using the previous example, we could have made a new file called "oxidases.faa" by doing the following:

```
extract_sequences "oxidase" myproteins.faa > oxidases.faa
```

2. Finding Prokka Results

You should locate your Prokka output directory. This was the directory that was created when you ran Prokka in the previous walkthrough. You should see files whose names begin with PROKKA in this directory.

3. Grepping tRNA

1. In the Prokka output folder (which we named `prokka_report` in the previous walkthrough), type `ls`.
2. You should see a PROKKA file that ends in `.ffn`
3. Try the following command:

```
grep "tRNA" *.ffn
```

4. This will return all headers with "tRNA" in the name.
5. If we're only interested in tRNA (and not proteins related to them), we can refine the search. Try:

```
grep "tRNA-...(…)" *.ffn
```

6. Now you should have a list of tRNA found in your organism.

4. Extracting tRNA

1. We can extract the tRNA that we found in the previous section by doing the following:

```
extract_sequence "tRNA-...(...)" *.ffn > trna.fa
```

2. This will extract the nucleotide sequences for each tRNA and put them in a file called **trna.fa**.
3. Confirm that the file exists using **ls** and look inside it using **less trna.fa**.

5. Extracting Ribosomal RNA

1. We can also extract ribosomal RNA by doing the following:

```
extract_sequences "ribosomal RNA" *.ffn > rna.fa
```

2. Type **ls** and confirm that the **rna.fa** file exists.
3. You can copy the sequences in rna.fa and paste them into NCBI blast.