COMP-8295
Course Project
MD Lutfar Rahman

Datasets:

Reference gnom length: 50000 - 100000
Dataset 1: 2 random reference gnomes
Dataset 2: 8 random reference gnomes
Dataset 3: 64 random reference gnomes

Reads:

Read length: 200 - 300
Read set 1: gnome1:1x, gnome2:2x
Read set 2: gnome1:2x, gnome2:4x
Read set 3: gnome1:3x, gnome2:5x

Precision and Sensitivity:

tp = number of correct classification
tn = number of wrong classification
fn = number of no classification

Precision = $\frac{tp}{tp + tn}$

Sensitivity = $\frac{tp}{tp + fn}$

Tasks:

1. I used the FMI code “fmi.py” given in the class.
   Code location: “code/fmi.py”

2. I used “random_dna()” from “util.py” given in the class. I tweaked it to my own needs. I wrote file saving code myself.
   code location: “random_dna()” in “code/util.py”

3. I wrote my own code. No one uses my code.
   code location: “generateRandomReads()” in “code/util.py”

4. I wrote my own code. No one uses my code.
   code location: “BuildIndexes()” and “AlignExatcly()” in “code/aligner.py”

5. I wrote my own code. No one uses my code.
   data location:
   Dataset 1: “code/data/dataset1/references/”
   Dataset 1 Reads: “code/data/dataset1/reads/”

   Dataset 2: “code/data/dataset2/references/”
   Dataset 2 Reads: “code/data/dataset2/reads/”

   Dataset 3: “code/data/dataset3/references/”
   Dataset 3 Reads: “code/data/dataset3/reads/”

   code location: “generateDataset()” and “generateReads()” in “code/datagenerator..py”

6. I wrote my own code. No one uses my code.
   code location: “runExperiment1()” in “code/main.py”
   my approach:
   ##
   ```
   seq = read to classify
   For each reference gnome g:
        freq = exact_align(g,seq)
        If freq > best_freq:
             Best_g = g
             Best_freq = feq
   Classify as Best_g
   ```
   ##

Report: Align exactly with 0 error:

	Runtime(sec)		Precision	Sensitivity
	Build index	query		
Dataset1	4	109	1.0	1.0
Dataset2	13	302	1.0	1.0
Dataset3	143	3802	1.0	1.0

7. and

8. I wrote my own code except “mutate_one_base()”. No one uses my code.
   code location: “mutate_one_base()”, “addErrors()” and “generateRandomReads()” in “code/util.py”

9. I wrote my own code. No one uses my code.
   code location: “runExperiment2()” in “code/main.py”

Report: Align exactly with 1% error:

	Runtime(sec)		Precision	Sensitivity
	Build index	query		
Dataset1	4	114	1.0	0.001693
Dataset2	13	316	1.0	0.001883
Dataset3	143	3916	1.0	0.000910

10. I wrote my own code. No one uses my code.
    code location: “AlignApproximately()” in “code/aligner.py”
    my approach: I slightly changed the idea and removed random index picking.
    ##
    ```
    matching_threshold = 8
    seq = read to classify
    For each reference gnome g:
          match,freq = psudo_align(g,seq)
          If match > 2*matching_threshold:
                Classify as g
          If match > matching_threshold:
                Best_match = match
                Best_g = g
    Classify as Best_g
    ```
    ##

11. I wrote my own code. No one uses my code.
    code location: “runExperiment3()” in “code/main.py”

Report: Align approximately with 1% error:

	Runtime(sec)		Precision	Sensitivity
	Build index	query		
Dataset1	4	104	.97	.95
Dataset2	13	298	.97	.94
Dataset3	143	3623	.96	.95

12. Haven't tried(Bonus)
13. Haven't tried(Bonus)

How to run:

1. "code/main.py" drives all the experiments.
2. "code/settings.py" has all the related settings