DNA

take sequence of DNA and match CSV

DNA → sequence of molecules - nucleotides

Each nucleotide:

either

adenine (A)

cytosine (C)

guanine (G)

thymine (T)

Background notes

Short Tandem Repeats (STR)

consecutive nucleotide structure

# of STR repeats varies

Multiple STRs improve accuracy

FBI uses 20 STRs

Specs

argv[1] → csv file w/ STRs

argv[2] → txt file w/ identity

re.search(____ , list[0])

STR

DNA Pseudocode

1. open CSV file / load into memory
2. open txt file / load into memory
3. Search txt file for each pattern
4. Record # of times pattern is found
5. Compare those numbers w/ numbers from CSV file, now in a list of dictionary entries
6. If complete match print name ~~at~~ the index of shortlen
7. If No match print no match

Note: if you're using a list variable containing string you still need to convert if you want to re.search(—, str(list))

step 4

loop

for key in shortlen[0] ← dict w/ str Rice keys
  if name – skip
  Find all matches of key in dna
  store in list (count)

Now iterate over count

Step 5 Compare values in Count w/ values in short.txt

loop
for i in range(len(count))

(Key){1, len(str(dna))}, dna

We need to find a method for isolating the longest sequence of repeated STRs and the number of repetitions

ABBAABBBAAABBBBAA

Saved repeat of B

result = 4

ind = x for x in range(len(dna)) if dna.startswith(key, x)

print(ind)

Rewrite over ind>