

Counting Point Mutations

Orr Shomroni

22 Oktober, 2021

Problem

Given two **strings** s and t of equal length, the **Hamming distance** between s and t , denoted $d_H(s, t)$, is the number of corresponding symbols that differ in s and t . See Figure 2.

Given: Two **DNA strings** s and t of equal length (not exceeding 1 kbp).

Return: The Hamming distance $d_H(s, t)$.

Sample Dataset

```
GAGCCTACTAACGGGAT
CATCGTAATGACGGCCT
```

Sample Output

7

```
def wrap(string):
    s=''
    for i in range(0,len(string),80):
        s+=string[i:i+80]
        s+='\n'
    return s

f=open("/home/orr/Dropbox/rosalind/bioinformatics_stronghold/rosalind_hamm.txt",'r')
s=f.readlines()
dH=0
for i in range(0,(len(s[0])-1),1):
    if s[0][i]!=s[1][i]:
        dH+=1

string="The Hamming distance between string s "+s[0].replace("\n","")+ " and string t "+s[1].replace("\n","")
print(wrap(string))

## The Hamming distance between string s ACTGCTTAAATGACATTCACTACCCAGCATCATTTTCTCGG
## CCGCAAATACGATGGGCTCTCTGGAAGGCAATTGCGTAGTGTTGACCCGGTCGCATTACACTCCGAGTGAACGAATACC
## CCTAGTCTACCACTTTGAATGGTGTATGATCAGTCAAAAGTAGTGGGATTGTTACCGTTAGCTTATCTTTTCGAGCTCCA
## CTCTATAACCACATGCTAGGCTATGTCAGACCCGGAGACGGGCCCTTAATTTTTACGTGAATCGACGACTCGGCATCTGT
## CCTAATAAGAAAAGAAAGCGGTCAGTTTAAGGTTTTATAGCTATACGCACGTTCTGGACTCCTTTTATCCATGCGAATCG
## CTACGAGCACTTTGACGCTCGTGGACTAATTTTCCTGTTCTGGAAGAATAGCACCCCTTACTTCATGAGTAGCTGCAG
## TTCGTGTCGGATATGGCGCACTCGTGGTCTGTAAGTACTCCTGGGGGTGTCGATAGCACCAGAGCAAGGACGCCTCAACC
## CGATCGTGGGTGACTTATGTAAAGACCCTGAGCCGGAACAAGGTCTGATCGTACTAATGGCGCTAAGATAGCACGAAAA
## TTTACGCCGAGGCAAAGAGGATTGAGAGTGCCCTGATCTGTGCGCCCTCTCAATGTATGCAGGCTATAGTCCTAATGTG
## AAATATAAGCCGTGGCGCATTGGGTGAGCGATCATAGGATAGCTCCCGTATATCATGAAATATCATTAACTTGACCT
```

```

## CGCTACCGGCTGCGGTTCAGCAGGGCCGACTGTTTCGCATTGCTGTGCTAAATCAAACGACTTTGGCCTGGGTAAACTGG
## ACCCCCCGTAACTGGTTGGTACTGACGCTCGACAGTTTGCTAGTCTGCCTCACTAAGCCCGTTACGTTCCACGCGCTCG
## CCTCGCATCCGGAGCGGGAATTAGATACCGTCCCAAGCACGACTACGTGCTGCGCATA and string t ACGGGTAG
## AATTACCCTCGAAGAGTCGCCGGACGGTTCTCCCTGGCCTAATTGATGCTAGCGGTGCAGGCCATTGCGGTTGGTAGTCA
## CTCTCTCACGTCCGGTTCGGAGCTGACGGATATCTATAAACTGCGTGATTGGCTTGTTCATGATCCAGGAAAAGGACTG
## TAATTGTCGCTTTGTGCTCCTCAATCCTGCGGAACGACAAACCAAGAATCTATGTATTTTCCCGCCAGCGCCATCCCT
## AATTCATATCAACAGAGAGGATTGAGCGTTTATCCTGGTACTATAAGAAACCGGCCAATTAATGGGTTTCATGACCATAC
## GTTTGTGGTGCTCTCGTGGTACCCATGCCACTCGCTACGAGGACTGAGAACTCTCGTCCGATGACGGTGCTATTCCCGGT
## TAAAGAGGGGACTTTTCTGTTTACGACCTCACGGGCGTTTCCGGTCTGGATAGCTGCTGTTATGCATTGATCCACTAT
## GTACTAAAACTCTTACTCGAGACGACAACTACCGAATGGGTGAAATCTTCCGGATGGTCTCTTATCCGGAACCCCTTGC
## TGACCGTCTGAGAGGCCCGGGGGGACAAATCAGTTTCCGTAAGTGGAACTAGAGAACAGGAGGCCACGACCGGGCGGAA
## CATAGCAAAACATGGAGGGCATACTCATCACCTGGTTTATAAGCCCTCGCGTTTTGCGTGGACGTACAGTCGGTTGTGGC
## CTTTAAATACAACTATCCCAATCACACGACCGTGCTTGGGGCCAGGTCCATCGCTCTCGTCTTAATGACTTCAAATA
## GGTAAC TAGAACGCCTACCGGCTCGGCAAACTGTACCCCTAGTCAGCAGGACCTTAGCCGCACTATACCTTGTGGGGGGC
## TCCCTCACAAGGCCTGTAAAGATCGACACCGCCATCCAGGAGACAGAGCTGGTTTCCGATACAATCCCAAGAGGGACAAT
## TACCTTCCCAT is 464

```