**(a)**

Given two sequences **x** and **y** as shown below

Determine theminimum number of edit operations (substitution, and indels) required to transform one into another

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**(b)**

Determine the Hamming distance between the strings: **CENTURY** and **SANCTUARY**

Determine the Levenshtein distance between the strings: **BIO-INFORMATICS** and **TRI-TELEMATICS**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Binary representation of a DNA sequence: Concept**

Sometimes DNA sequence analysis can be done by converting the sequence into binary format. Foe example, suppose the following *dibit* representation is pursued: A ⇒ (11); C ⇒ (01); G ⇒ (10); T ⇒ (00). The a sequence ACCTGCA, for example can be written as: 11 01 01 00 10 01 11

**(c)**

By constructing binary format of the pair of sequences x and y given **below** determine the Hamming distance between them.

(*Hint*: For **binary strings** *a* and *b,* the Hamming distance is equal to the number of ones resulting in *a* XOR *b* operation).

**(d)**

Given a template sequence: CCCAAGGGGTTCCAATG. Identify the underlying

mutations and derivatives namely, point-mutations, deletions, inversions, transportations, duplications, insertions in the following set of strings that resemble the template:

CCCAAGGGGTTTCAATG

CCCAAGGGGTTTCxxxx

CCGGAACGGTTTC

TTTCCCGGAACGG

TTTCCCGGGGAAGG

TTTCCCGGTTAACTTTGG

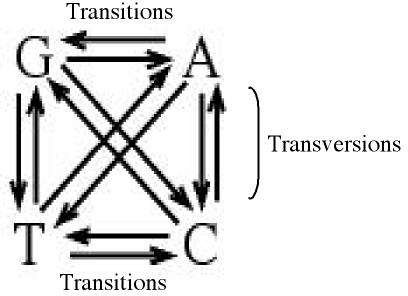
TTTCCCGGTTAACTTGG

How will you designate the following sequence in relation to the template?

AAAGGCCAATTGAAACC

**(e)**

Transition mutations are more common than transversions mutations.



Construct a matrix to illustrate such characteristics of the mutations.Assume proportionate percentage to depict each type of mutation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **A** | **T** | **G** | **C** |
| **A** |  |  |  |  |
| **T** |  |  |  |  |
| **G** |  |  |  |  |
| **C** |  |  |  |  |

x: **T** **A G C T A T C G G G A A C T G**

y **G C T C A C G G T T G G G A C T**