Graphical user interface, text, application, email

Description automatically generatedTo create a substitution matrix that reflects the lower frequency of transition mutations, we can use the following values:

A G T C A 1 -1 -3 -1 G -1 1 -1 -3 T -3 -1 1 -1 C -1 -3 -1 1

Here, the transition mutations have a penalty of -1, while the transversion mutations have a penalty of -3. This results in a higher penalty for transversions, reflecting their lower frequency.

To reflect the lower frequency of transition mutations compared to transversions, we can modify the simplified substitution matrix by assigning a lower score for transition mutations and a higher score for transversions.

One possible substitution matrix that reflects this is:

A G T C

A 1 -2 -1 -2

G -2 1 -2 -1

T -1 -2 1 -2

C -2 -1 -2 1

In this matrix, a transition mutation receives a score of -1 (compared to -2 in the original matrix) and a transversion mutation receives a score of -2 (compared to -1 in the original matrix). The diagonal elements remain the same as in the original matrix.

Note that there can be multiple possible substitution matrices that reflect the lower frequency of transition mutations. The specific choice of matrix can depend on factors such as the specific context of the alignment problem and the goals of the analysis.

The above code is a Python implementation of a substitution matrix with scores for DNA sequence alignments. Here is a brief explanation of the code:

1. The NumPy library is imported.
2. Four scores are defined: match score (1), mismatch score (-1), transition score (-3), and transversion score (-1).
3. The four nucleotides are defined as a list: A, G, T, C.
4. A 4x4 substitution matrix is initialized with zeros using the NumPy **zeros** function.
5. A nested loop is used to fill in the substitution matrix with scores. If the row and column indices are equal, the match score is used. Otherwise, if the pair of nucleotides represents a transition (A<->G or T<->C), the transition score is used. Otherwise, the transversion score is used.
6. The substitution matrix is converted to integers using the **astype** method.
7. The substitution matrix is printed to the console.

The resulting substitution matrix can be used to score alignments of DNA sequences based on how well the nucleotides match or differ.

6331 5349