**PRACTICAL ON SEQUENCE ALIGNMENTS FUNDAMENTALS**

**STUDENT: PIERRE WENSEL**

**Q1**: Based on this multiple sequence alignment, construct a substitution matrix. Include all Tables (frequency of occurrence, expected frequency of occurrence, Observed frequency of occurrence)

ACAG

ACTG

ATAC

CTAC

GTAC

ACAT

**TABLE1:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Nucleotide** | **Frequency of Occurrence** |  | **Nucleotide Pair** | **Expected Frequency of Occurrence (E )** | **Expected Frequency of Occurrence (E )** |
| A | 9 |  | AA | (9/24)\*(9/24) | 0.140625 |
| G | 3 |  | AG | (9/24)\*(3/24) | 0.046875 |
| T | 5 |  | AT | (9/24)\*(5/24) | 0.078125 |
| C | 7 |  | AC | (9/24)\*(7/24) | 0.109375 |
|  |  |  | GG | (3/24)\*(3/24) | 0.015625 |
|  |  |  | GT | (3/24)\*(5/24) | 0.026041667 |
|  |  |  | GC | (3/24)\*(7/24) | 0.036458333 |
|  |  |  | TT | (5/24)\*(5/24) | 0.043402778 |
|  |  |  | TC | (5/24)\*(7/24) | 0.060763889 |
|  |  |  | CC | (7/24)\*(7/24) | 0.085069444 |
| TOTAL | 4\*6=24 |  |  |  |  |

**TABLE 2:**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Nucleotide Pair** | **Frequency of Occurrence** | | | **Frequency of Occurrence** | | | **Observed Frequency of Occurrence (O)** | | **Observed Frequency of Occurrence (O)** | **OddsRatio(O/E)** | **2log2(O/E)** |
| AA | (3+2+1)+(4+3+2+1) | | | 16 | | | (16/57) | | 0.280701754 | 1.9961014 | 1.99437 |
| AG | 4 | | | 4 | | | (4/57) | | 0.070175439 | 1.497076 | 1.164295 |
| AT | 5 | | | 5 | | | (5/57) | | 0.087719298 | 1.122807 | 0.33422 |
| AC | 4 | | | 4 | | | (4/57) | | 0.070175439 | 0.641604 | -1.28049 |
| GG | 1 | | | 1 | | | (1/57) | | 0.01754386 | 1.122807 | 0.33422 |
| GT | 2 | | | 2 | | | (2/57) | | 0.035087719 | 1.3473684 | 0.8602888 |
| GC | 1+3+3 | | | 7 | | | (7/57) | | 0.122807018 | 3.3684211 | 3.504145 |
| TT | 2+1 | | | 3 | | | (3/57) | | 0.052631579 | 1.2126316 | 0.5562826 |
| TC | 3+3+3 | | | 9 | | | (9/57) | | 0.157894737 | 2.5984962 | 2.7553539 |
| CC | (2+1)+(2+1) | | | 6 | | | (6/57) | | 0.105263158 | 1.2373792 | 0.6145753 |
| TOTAL |  | | | 57 | | |  | |  |  |  |
| **Substitution Matrix** | | **A** | **T** | | **C** | **G** | |
| **A** | | 1.99 |  | |  |  | |
| **T** | | 0.33 | 0.56 | |  |  | |
| **C** | | -1.28 | 2.76 | | 0.61 |  | |
| **G** | | 1.16 | 0.86 | | 3.50 | 0.33 | |

**Substitution matrix (ABOVE)**

**Q2**: Which one of the three alignments is the best for ACGGA and ACCAGA Sequences according to the previous substitution matrix? Score the three pairwise sequence alignments. Gap penalty:-2

ACGG-A

ACCAGA

SCORE=7.25

A-CGGA

ACCAGA

SCORE=4.08

ACG-GA

ACCAGA

SCORE=6.42

**Score1=**1.99+0.61+3.50+1.16-2+1.99=**7.25**

**Score2=**1.99-2+0.61+1.16+0.33+1.99=**4.08**

**Score3=**1.99+0.61+3.50-2+0.33+1.99=**6.42**

**Best Sequence Alignment=Sequence Alignment # 1**

**Q3**:Use BLOSUM62, a gap penalty of -2, and a gap extension of -1 to score this pairwise sequence alignment:

VL-TGTT-S

VLMTAT-NS

**Score=4+4+(-2)+5+0+5+(-2)+(-2)+4=16**

**Sequence identity= # Aligned residues with identical characteristics/Length of shorter of the two sequences \*100%=(5/7)\*100=71.42%**