Feng-Doolittle Unit Tests

Test 1 (Hint: Notation from original paper! Formula was wrong! Mismatches were not considered!)
Input

Sequence a: ACGT
Sequence b: AT
Sequence c: GCC

Gap opening: 0 (easifies later visual proofment)
Enlargement: -2
Match: 1 (and 0 for placeholder #)
Mismatch: -1

Output (Pairwise Alignment)

	Alignment-Length	Gaps	Gap-starts	Score
(a,b)	4	2	1	-2
(a,c)	4	1	1	-3
(b,c)	3	1	1	-4

```
Seq1  ACGT
      *  *
Seq2  A__T


Seq1  ACGT
      |* |
Seq2  GC_C


Seq1  _AT
       ||
Seq2  GCC
```

Hint: More alignments exists, but only one is computed!

Output (Distances)
$S_{a,b}^{rand}$

$$= \frac{1}{4}\begin{pmatrix} s(A_a, A_b) \cdot N_A(a) \cdot N_A(b) + s(A_a, T_a) \cdot N_A(a) \cdot N_T(b) \\ + s(C_a, A_b) \cdot N_C(a) \cdot N_A(b) + s(C_a, T_a) \cdot N_C(a) \cdot N_T(b) \\ + s(G_a, A_b) \cdot N_G(a) \cdot N_A(b) + s(G_a, T_b) \cdot N_G(a) \cdot N_T(b) \\ + s(T_a, A_b) \cdot N_T(a) \cdot N_A(b) + s(T_a, T_b) \cdot N_T(a) \cdot N_T(b) \end{pmatrix}$$

$$+ 2 \cdot enlarge$$

$$= \frac{1}{4}\begin{pmatrix} 1 + (-1) + (-1) + (-1) \\ + \\ (-1) + (-1) + (-1) + 1 \end{pmatrix} + 2 \cdot (-2) = \frac{-4}{4} - 4 = -5$$

$$S_{a,b}^{max} = \frac{4+2}{2} = 3$$

$$S_{a,b}^{eff} = \frac{S(a,b) - S_{a,b}^{rand}}{S_{a,b}^{max} - S_{a,b}^{rand}} = \frac{-2 - (-5)}{3 - (-5)} = \frac{3}{8}$$

$$D(a,b) = -\ln S_{a,b}^{eff} \approx 0.98 \approx 1$$

$$S_{a,c}^{rand} = \frac{1}{4} \begin{pmatrix} s(A_a, G_b) \cdot N_A(a) \cdot N_G(b) + s(A_a, C_b) \cdot N_A(a) \cdot N_C(b) \\ +s(C_a, G_b) \cdot N_C(a) \cdot N_G(b) + s(C_a, C_b) \cdot N_C(a) \cdot N_C(b) \\ +s(G_a, G_b) \cdot N_G(a) \cdot N_G(b) + s(G_a, C_b) \cdot N_G(a) \cdot N_C(b) \\ +s(T_a, G_b) \cdot N_T(a) \cdot N_G(b) + s(T_a, C_b) \cdot N_T(a) \cdot N_C(b) \end{pmatrix} + 1 \cdot enlarge$$

$$= \frac{1}{4} \begin{pmatrix} -1 + (-2) + (-1) + 2 \\ + \\ 1 + (-2) + (-1) + (-2) \end{pmatrix} + 1 \cdot (-2) = \frac{-6}{4} - 2 = -3.5$$

$$S_{a,c}^{max} = \frac{4 + 3}{2} = 3.5$$

$$S_{a,c}^{eff} = \frac{S(a,c) - S_{a,c}^{rand}}{S_{a,c}^{max} - S_{a,c}^{rand}} = \frac{-3 - (-3.5)}{3.5 - (-3.5)} = \frac{0.5}{7}$$

$$D(a,c) = -\ln\left(S_{a,c}^{eff}\right) \approx 2.639 \approx 3$$

```
Seq1   _AT
        ||
Seq2   GCC
```

$$S_{b,c}^{rand} = \frac{1}{3} \cdot \begin{pmatrix} s(A_a, G_b) \cdot N_A(a) \cdot N_G(b) + s(A_a, C_b) \cdot N_A(a) \cdot N_C(b) \\ +s(T_a, G_b) \cdot N_T(a) \cdot N_G(b) + s(T_a, C_b) \cdot N_T(a) \cdot N_C(b) \end{pmatrix} + 1 \cdot enlarge$$

$$= \frac{1}{3} \cdot \left(-1 + (-2) + (-1) + (-2)\right) - 2 = \frac{-6}{3} - 2 = -4$$

$$S_{b,c}^{max} = \frac{2 + 3}{2} = 2.5$$

$$S_{b,c}^{eff} = \frac{S(b,c) - S_{b,c}^{rand}}{S_{b,c}^{max} - S_{b,c}^{rand}} = \frac{-4 - (-4)}{2.5 - (-4)} = \frac{0}{6.5} \leq 0 \rightarrow S_{a,c}^{eff} = \frac{0.001}{6.5} = \frac{1}{6500}$$

$$D(b,c) = -\ln\left(S_{b,c}^{eff}\right) \approx 8.78 \approx 9$$

```
Seq1   _AT
        ||
Seq2   GCC
```

Output (Phylogenetic Tree)

1.

$d_{min} = 1$

	a	b	c
a	0	1	3
b		0	9
c			0

2.

$$\mathcal{C} = \big((\mathcal{C} - \{a\}) - \{b\}\big) \cup \{d\}$$

	a	b	c	d
a	0	1	3	
b		0	9	
c			0	6
d				0

3.

$$dist(d, a) = dist(d, b) = \frac{1}{2} = 0.5$$

4.
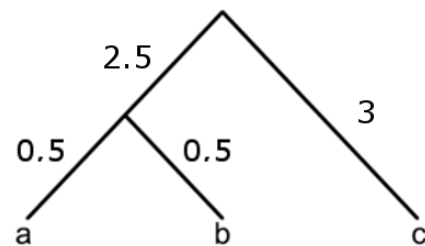
$$dist(c, d = \{a, b\}) = \frac{|a| \cdot dist(c, a) + |b| \cdot dist(c, b)}{|a| + |b|} = \frac{1 \cdot 3 + 1 \cdot 9}{1 + 1} = 6$$

----

1) $d_{min} = 6$

2) $\mathcal{C} = \big((\mathcal{C} - \{c\}) - \{d\}\big) \cup \{e\}$

3) $dist(e, c) = dist(e, d) = \frac{d_{min}}{2} = 3$

	a	b	c	d	e
a	0	1	3		
b		0	9		
c			0	6	
d				0	
e					0

Output (Joinment)
1.
ACGT
A##T

2.
ACGT   GCC
A##T   and

Seq1 ACGT   Seq1 ACGT   Seq1 A##T   Seq1 A##T
      |* | |*| | |
Seq2 GC_C   Seq2 GCC_   Seq2 GCC_   Seq2 _GCC
Score -3   **Score** -3   **Score** -3   **Score** -3

		A	#	#	T
	0	-2	-2	-2	-4
G	-2	-1	-2	-2	-3
C	-4	-3	-1	-2	-3
C	-6	-6	-5	-3	-3

Output (Final)
GC_C
ACGT
A__T
SoP-Score -9