Feng-Doolittle Unit Tests

```
Hint:   Many test values are taken from project Algorithms
        for Bioninformatics of Alexander Mattheis or the lectures.
```

Test 1 (Hint: Notation from original paper used!)

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)

$$S_{a,b}^{rand} = \frac{1}{4}\begin{pmatrix} s(A_a, A_b)\cdot N_A(a)\cdot N_A(b) + s(A_a, A_a)\cdot N_A(a)\cdot N_T(b) \\ +s(T_a, T_b)\cdot N_T(a)\cdot N_A(b) + s(T_a, T_a)\cdot N_T(a)\cdot N_T(b) \end{pmatrix} + 2\cdot enlarge$$

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

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

$$S_{a,b}^{eff} = \frac{S(a,b) - S_{a,b}^{rand}}{S_{a,b}^{max} - S_{a,b}^{rand}} = \frac{-2-(-3)}{3-(-3)} = \frac{1}{6}$$

$$D(a,b) = -\ln S_{a,b}^{eff} \approx 1.792 \approx 2$$

$$S_{a,c}^{rand} = \frac{1}{4}\begin{pmatrix} s(C_a, C_b)\cdot N_C(a)\cdot N_C(b) + s(C_a, C_b)\cdot N_C(a)\cdot N_G(b) \\ +s(G_a, G_b)\cdot N_G(a)\cdot N_G(b) + s(G_a, G_b)\cdot N_G(a)\cdot N_C(b) \end{pmatrix} + 1\cdot enlarge$$
$$= \frac{1}{4}(2+1+1+2) + 1\cdot(-2) = 1.5 - 2 = -0.5$$

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

$$S_{a,c}^{eff} = \frac{S(a,c) - S_{a,c}^{rand}}{S_{a,c}^{max} - S_{a,c}^{rand}} = \frac{-3-(-0.5)}{3.5-(-0.5)} = \frac{-2.5}{4} < 0 \rightarrow S_{a,c}^{eff} = \frac{0.001}{4} = \frac{1}{4000}$$

$$D(a,c) = -\ln\left(S_{a,c}^{eff}\right) \approx 8.394 \approx 8$$

$$S_{b,c}^{rand} = \frac{1}{3}\cdot 0 + 1\cdot enlarge$$
$$= -2$$

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

$$S_{b,c}^{eff} = \frac{S(b,c) - S_{b,c}^{rand}}{S_{b,c}^{max} - S_{b,c}^{rand}} = \frac{-4-(-2)}{2.5-(-2)} = \frac{-2}{4.5} < 0 \rightarrow S_{a,c}^{eff} = \frac{0.001}{4.5} = \frac{1}{4500}$$

$$D(b,c) = -\ln\left(S_{b,c}^{eff}\right) \approx 8.412 \approx 8$$

Output (Phylogenetic Tree)

1.

$d_{min} = 2$

	a	b	c
a	0	2	8
b		0	8
c			0

2.

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

	a	b	c	d
a	0	2	8	
b		0	8	
c			0	8
d				0

3.

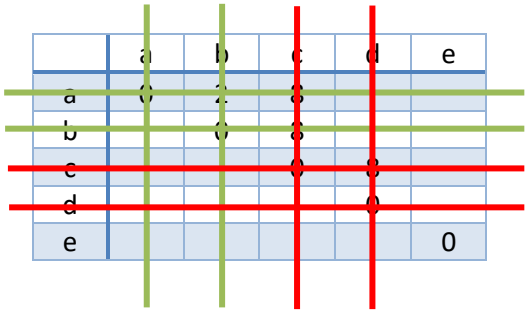
$dist(d,a) = dist(d,b) = \frac{2}{2} = 1$

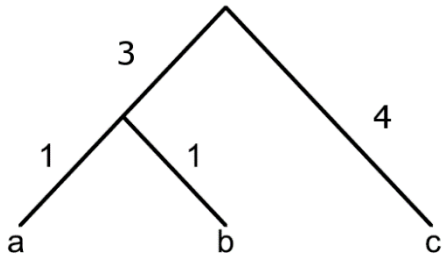
4.

$$dist(c, d = \{a, b\}) = \frac{|a| \cdot dist(c, a) + |b| \cdot dist(c, b)}{|a| + |b|} = \frac{1 \cdot 8 + 1 \cdot 8}{1 + 1} = 8$$

---

1) $d_{min} = 8$
2) $\mathcal{C} = \big((\mathcal{C} - \{c\}) - \{d\}\big) \cup \{e\}$
3) $dist(e, c) = dist(e, d) = \frac{d_{min}}{2} = 4$





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)

```
ACGT
A__T
GCC_
```