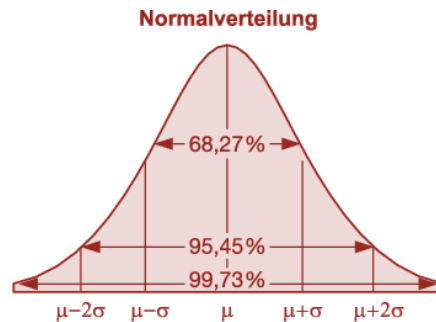


THE . RETORT . COUR-TEOUS-  
THE . REPLY- . CHURL---ISH

4.  $P$  ist die Wahrscheinlichkeit, dass ein gewisser Scorewert zufällig übertroffen wird. Welchem Wert von  $P$  würde ein Z-Score von 1 in einer Normalverteilung entsprechen?

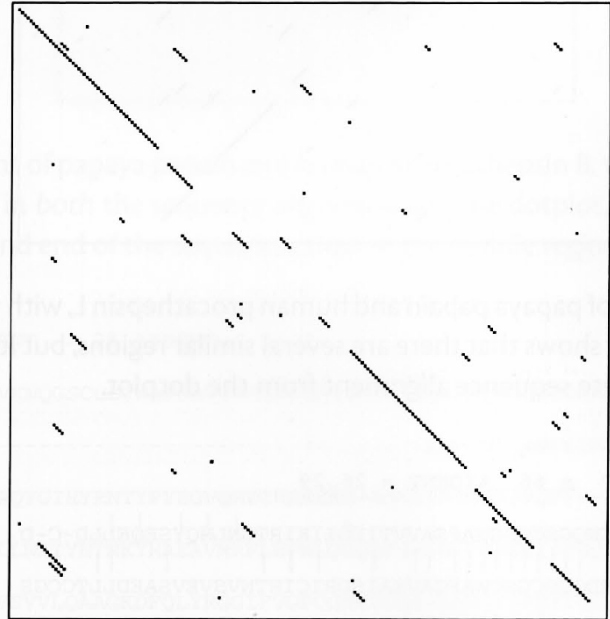


$(100 - 68.27)/2 = 15.865\%$  liegen über  $+1\sigma$  (Standardabweichung)

$$P(\text{Z-Score} > 1) = \int_1^{\infty} \mu_N(x) dx = \int_1^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \approx 0.16$$

5. Die unten stehende Figur zeigt das Sequenz Alignment von Papaya Papain mit Kiwi Actinidin und den entsprechenden Dotplot. Das Sequenz Alignment zeigt zwei Orte, wo mehrere Aminosäurereste aus der Papain Sequenz gestrichen werden. Markieren Sie im Dotplot die ungefähren Positionen dieser Deletionen.

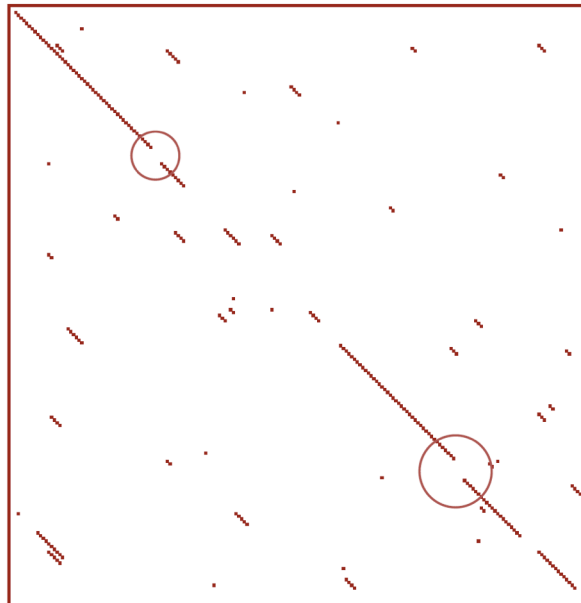
## PAPA\_CARPA / ACTN\_ACTCH



NPOS = 219 NIDENT = 102 %IDENT = 46.58

```

IPEYVDRQKGAVTPVKNQSGSCWAFSAVVTIEGIIKIRTGNLNQYSEQELLDCDR--
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
LPSYVDWRSAGAVVDIKSQGECGGCWAFSAIATVBGINKITSGSLISLSEQELIDCGRTQ
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RSYGCNGGYPWSALQ-LVAQYGIHYRNTYPYEGVQRYCRSREKGPYAAKTDGVRQVQPYN
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
NTRGCDGGYITDGFQFIINDGGINTEENYPYTAQDGDVALQDKYVTDITYENVYPYNN
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QGALLYSIANQPVSVVLQAAGKDFQLYRGGIFVPGPCGNKVDHAVAAGVGP---NYILI
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
EWALQTAVTYQPVSVLDAAGDAFKQYASGIFTGPGCTAVDHAIVIVGYGTEGGVDYWIV
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
KNSWGTGWGENGYIRIKRGTGNSYGVCGLYTSSFYPVKN
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
KNSWDTTWGEEGYMRILRNVGGA-GTCGIATMPSYPVKY
  
```



6. Vergleiche der DNA Sequenzen homologer Chromosomen von verschiedenen Personen zeigen, dass im Mittel eines von 700 Basenpaaren in nicht-kodierender DNA unterschiedlich ist. 95 % des menschlichen Genoms ist nicht-kodierend. Schätzen Sie die Zahl der Polymorphismen im menschlichen Genom, um eine Abschätzung der Zahl potentieller DNA Marker zu erhalten.

$$(\text{Size of human genome}) \times (\text{fraction of non-coding DNA}) / (\text{base pairs per polymorphism}) = 3 \times 10^9 \times 0.95 / 700 \approx 4 \times 10^6$$

7. Zeigen Sie die Rechnungen, die im folgenden Beispiel zur Anwendung des Needleman- Wunsch Algorithmus zur dynamischen Programmierung mit Score 0 für Match, 20 für Mismatch und 25 für Insertion oder Deletion zum Eintrag „65“ führen. Warum gehen von diesem Eintrag zwei Pfeile aus?

	ϕ	a	t	g
ϕ	0	←25	←50	←75
g	25	20	←45	50
g	50	45	40	45
a	75	50	65	60
a	100	75	70	85
t	125	100	75	90
g	150	125	100	75
g	175	150	125	100

3 mögliche Wege zum Feld mit „65“: von links  $50 + 25 = 75$  (Deletion), diagonal  $45 + 20 = 65$  (Mismatch); von oben  $40 + 25 = 65$  (Deletion) → Minimum = 65 kann diagonal oder von oben erreicht werden.

8. Bestimmen Sie das optimale globale Alignment der beiden Proteinsequenzen AWTYA und QWERTY mit Hilfe des Needleman-Wunsch Alignment Algorithmus unter Verwendung der BLOSUM62 Scoring Matrix (vgl. Übung 2b) und einer Strafe für Lücken (gap penalty) von 10.

a) Durch manuelle dynamische Programmierung. Wie gross ist der Score?

	ϕ		A		W		T		Y		A
ϕ	0	←	-10	←	-20	←	-30	←	-40	←	-50
Q	-10	↑	-1		-11	←	-21	←	-31	←	-41
W	-20	↑	-11	↑	10	←	0	←	-10	←	-20
E	-30	↑	-21	↑	0	←	9	←	-1	←	-11
R	-40	↑	-31	↑	-10	←	-1	←	7	←	-2
T	-50	↑	-40	↑	-20	←	-5	←	-3	←	7
Y	-60	↑	-50	↑	-30	←	-15	←	2	←	-3

AWTYA-  
QWERTY

Score = -3

b) Mit Hilfe des Webservers [http://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](http://www.ebi.ac.uk/Tools/psa/emboss_needle/).

```

=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 10.0
#
# Length: 6
# Identity:      1/6 (16.7%)
# Similarity:   1/6 (16.7%)
# Gaps:         1/6 (16.7%)
# Score: -3.0
#
#
=====

EMBOSS_001      1 AWTYA-      5
                  .|...
EMBOSS_001      1 QWERTY      6

```

9. Wiederholen Sie Übung 8 mit einer Strafe für Lücken von 1.

a) Durch manuelle dynamische Programmierung. Wie gross ist der Score?

	Φ		A		W		T		Y		A
Φ	0	←	-1	←	-2	←	-3	←	-4	←	-5
	↑	↖				↖		↖		↖	
Q	-1		-1	←	-2	←	-3	←	-4	←	-5
	↑		↑	↖							
W	-2		-2		10	←	9	←	8	←	7
	↑	↖	↑		↑	↖				↖	
E	-3		-3		9		9	←	8	←	7
	↑	↖	↑		↑	↖	↑	↖	↑	↖	
R	-4		-4		8		8	←	7		7
	↑	↖			↑	↖					
T	-5		-4		7		13	←	12	←	11
	↑		↑		↑		↑	↖			
Y	-6		-5		6		12		20	←	19

AW--TYA

QWERTY-

Score = 19

b) Mit Hilfe des Webservers [http://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](http://www.ebi.ac.uk/Tools/psa/emboss_needle/).

```

=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 1.0
# Extend_penalty: 1.0
#
# Length: 7
# Identity:      3/7 (42.9%)
# Similarity:   3/7 (42.9%)
# Gaps:         3/7 (42.9%)
# Score: 19.0
#
#
=====

EMBOSS_001      1 AW--TYA      5
                  .|  ||
EMBOSS_001      1 QWERTY-     6

```

8'. Wiederholen Sie Übung 8 mit lokalem Alignment (Smith-Waterman Algorithmus).  
a) Durch manuelle dynamische Programmierung. Wie gross ist der Score?

Zusätzliche Regeln für lokales Alignment:

- Score in einem Feld immer  $\geq 0$
- Keine Pfeile eintragen, die zu 0 führen.
- Lokales Alignment beim grössten Scorewert in der Matrix beginnen.
- So weit wie möglich den Pfeilen entlang zurückgehen (d.h. bis ein 0 kommt).

	Φ	A	W		T		Y		A
Φ	0	0	0		0		0		0
Q	0	0	0		0		0		0
W	0	0	11	←	1		2		0
			↑	↖				↖	
E	0	0	1		10		0		1
						↖			
R	0	0	0		0		8		0
								↖	
T	0	0	0		5		0		8
						↖			
Y	0	0	2		0		12	←	2

TY

TY

Score = 12

b) Mit Hilfe des Webservers [http://www.ebi.ac.uk/Tools/psa/emboss\\_water/](http://www.ebi.ac.uk/Tools/psa/emboss_water/).

```
#####
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 10.0
#
# Length: 2
# Identity:      2/2 (100.0%)
# Similarity:    2/2 (100.0%)
# Gaps:          0/2 ( 0.0%)
# Score: 12.0
#
#
#####

EMBOSS_001      3 TY      4
                 ||
EMBOSS_001      5 TY      6
```

9'. Wiederholen Sie Übung 9 mit lokalem Alignment (Smith-Waterman Algorithmus).  
a) Durch manuelle dynamische Programmierung. Wie gross ist der Score?

	Φ	A	W		T		Y		A
Φ	0	0	0		0		0		0
Q	0	0	0		0		0		0
W	0	0	11	←	10	←	9	←	8
			↑	↖				↖	
E	0	0	10		10	←	9	←	8
			↑	↖	↑	↖	↑	↖	
R	0	0	9		9	←	8		8
			↑	↖					
T	0	0	8		14	←	13	←	12
			↑		↑	↖			
Y	0	0	7		13		21	←	20

W--TY

WERTY

Score = 21

b) Mit Hilfe des Webservers [http://www.ebi.ac.uk/Tools/psa/emboss\\_water/](http://www.ebi.ac.uk/Tools/psa/emboss_water/).

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 1.0
# Extend_penalty: 1.0
#
# Length: 5
# Identity:      3/5 (60.0%)
# Similarity:   3/5 (60.0%)
# Gaps:         2/5 (40.0%)
# Score: 21.0
#
#=====

EMBOSS_001      2 W--TY      4
                  |  ||
EMBOSS_001      2 WERTY      6
```

10. Finden Sie die Sequenz der dUTPase des Fowl Adenovirus 1 in der Proteinsequenz Datenbank <http://www.ncbi.nlm.nih.gov/protein/> und stellen Sie sie im FASTA Format dar.

<http://www.ncbi.nlm.nih.gov/protein/Q89662.1>

```
>gi|7531092|sp|Q89662.1|DUT_ADEG1 RecName: Full=Deoxyuridine 5'-triphosphate
nucleotidohydrolase; Short=dUTPase; AltName: Full=dUTP pyrophosphatase
MDPFGSSSVPPCSTSDLPEPKLYFVRLSPHAVPPVRATHGAAGYDLFSAYDIKVPARGRALVPTDLVFQF
PPGCYGRIAPRSLAAKFFIDVGAGVIDPDYRGNVSVVLFNFSESSFNIRRGDRVAQLILERIMVPELSE
LTQLGETDRGASGFGSTGMGAVDRNQRSVLEWLTPGSR
```

11. Suchen Sie mit Hilfe des BLAST Servers <http://blast.ncbi.nlm.nih.gov/> andere Proteine mit ähnlicher Sequenz wie die dUTPase des Fowl Adenovirus 1 aus der vorangehenden Übung.

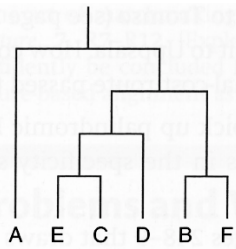
a) Führen Sie die Suche mit den BLAST Standard Parametern durch.

b) Finden Sie das am nächsten verwandte menschliche Protein.

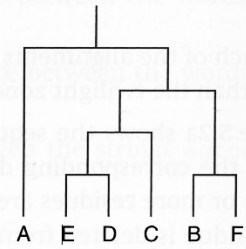
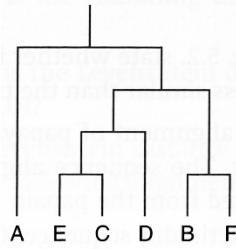
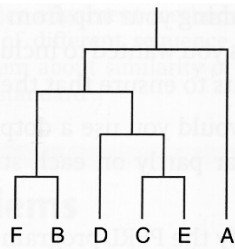
c) Wiederholen Sie die Suche, indem Sie die BLOSUM250 Scoring Matrix verwenden.

12. Haben die folgenden Paare von Bäumen identische Topologie?

a)



b)



a) Ja

b) Nein

13. Führen Sie ein Clustering der Cytochrome c Sequenzen verschiedener Spezies mit folgenden Abständen durch

	Turtle A	Man B	Tuna C	Chicken D	Moth E	Monkey F	Dog G
Turtle							
Man	19						
Tuna	27	31					
Chicken	8	18	26				
Moth	33	36	41	31			
Monkey	18	1	32	17	35		
Dog	13	13	29	14	28	12	

a) Verwenden Sie dazu die Unweighted Pair-Group Method with Arithmetic mean (UPGMA), bei der der Abstand zwischen zwei Clustern gleich dem arithmetischen Mittel aller Abstände zwischen Einzelsequenzen in den beiden Clustern ist.

b) Zeichnen Sie den zugehörigen phylogenetischen Baum, so dass die gesamte Kantenlänge bis zu einem Branch dem halbem Abstand zwischen den entsprechenden Clustern entspricht.

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

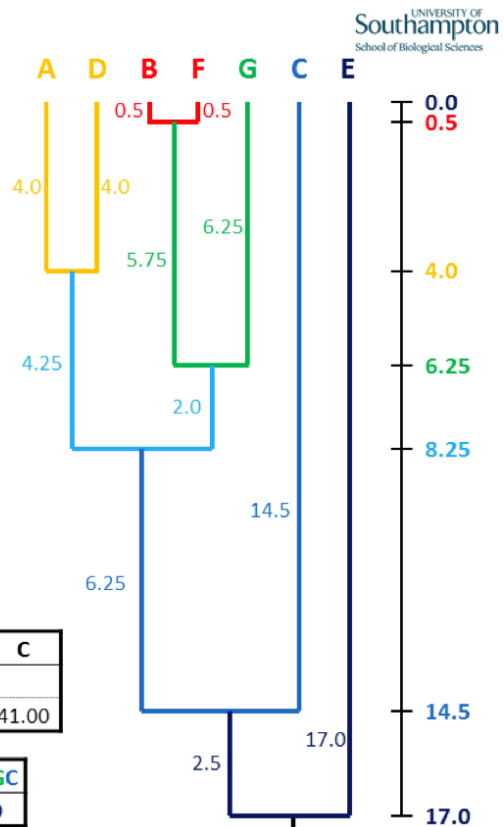
	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	8.00	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	AD	BF	C	E
BF	18.00			
C	26.50	31.50		
E	32.00	35.50	41.00	
G	13.50	12.50	29.00	28.00

	AD	BFG	C
BFG	16.50		
C	26.50	30.67	
E	32.00	33.00	41.00

	ADBFG	C
C	29.00	
E	32.60	41.00

	ADBFGC
E	34.00



UNIVERSITY OF  
Southampton  
School of Biological Sciences