

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?

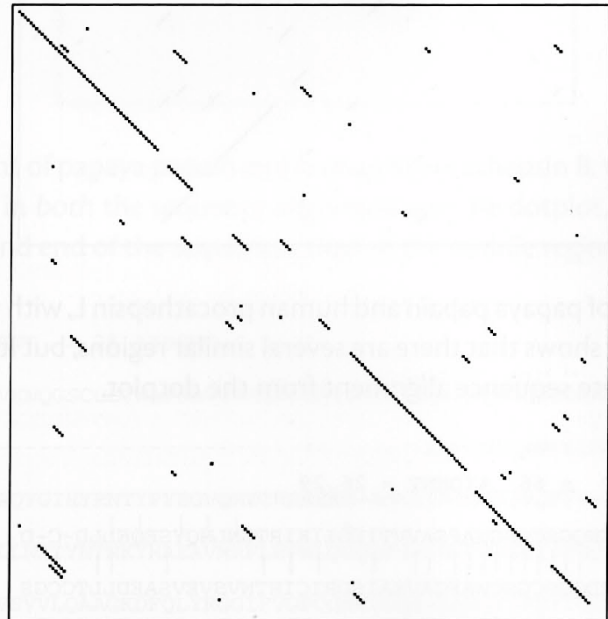
5. die unten stehende Figur zeigt das Sequenz Alignment von Papaya Papain mit Kiwi Actinidin und den entsprechenden Dotplot. Das Sequenz Alignment zeigt zwei Orte, wo mehrere Aminosäurereste aus der Papain Sequenz gestrichen werden. Markieren Sie im Dotplot die ungefähren Positionen dieser Deletionen.

```

NPOS = 219 NIDENT = 102 %IDENT = 46.58
IPEYVDWRQKGAVTTPVKNQSGSCGWAFSAVVTIEGIIKIRTGNLNQYSEQELLDCCR--
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
LPSYVDWRSAGAVVDIKSQGECGCWAFSAIATVEGINKITSGSLISLSEQELIDCGRTQ
RSYGCNGGGYPWSALQ-LVAQYGIHYRNTYPYEGVQRYCRSREKGPYAAKTDGVRVQVPYN
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
NTRGCDGGYITDGFQFIINDGGINTEENYPYTAQDGDCDVALQDQKYYVTIDTYENVPYNN
QGALLYSIANQPVSVVLQAAGKDFQLYRGGIFVPGCGNKVDHAVAAGVGYP---NYILI
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
EWALQTAVTYQPVSVLDAAGDAFKQYASGIFTGPCGTAVDHAIVIVGYGTEGGVDYWIV
KNSWGTGWGENGYIRIKRGTGNSYGVCGLYTSSFPVKN
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
KNSWDTTWGEEGYMRILRNVGGA-GTCGIATMPSYPVKY

```

PAPA_CARPA / ACTN_ACTCH



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.

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 | ↑ | ↖ | ↖ | ↖ |
| g | 25 | 20 | ←45 | 50 |
| g | ↑ | ↖ | ↖ | ↖ |
| a | 50 | 45 | 40 | 45 |
| a | ↑ | ↖ | ↖ | ↖ |
| a | 75 | 50 | 65 | 60 |
| a | ↑ | ↖ | ↖ | ↖ |
| t | 100 | 75 | 70 | 85 |
| t | ↑ | ↑ | ↖ | ↖ |
| t | 125 | 100 | 75 | 90 |
| t | ↑ | ↑ | ↑ | ↖ |
| g | 150 | 125 | 100 | 75 |
| g | ↑ | ↑ | ↑ | ↖ |
| g | 175 | 150 | 125 | 100 |

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.

- Durch manuelle dynamische Programmierung. Wie gross ist der Score?
- Mit Hilfe des Webservers http://www.ebi.ac.uk/Tools/psa/emboss_needle/.

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

- Durch manuelle dynamische Programmierung. Wie gross ist der Score?
- Mit Hilfe des Webservers http://www.ebi.ac.uk/Tools/psa/emboss_needle/.

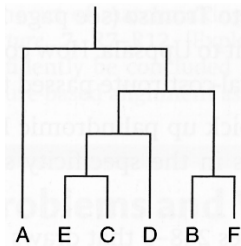
10. Finden Sie die Sequenz der dUTPase des Fowl Adenovirus 1 in der Proteinsequenz Datenbank <http://www.ncbi.nlm.nih.gov/protein/> und stellen Sie sie im FASTA Format dar.

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.

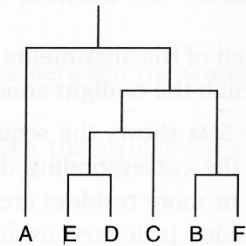
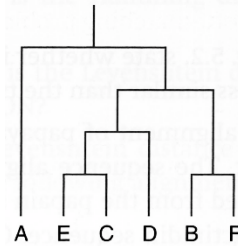
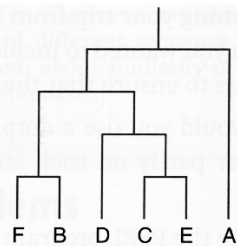
- Führen Sie die Suche mit den BLAST Standard Parametern durch.
- Finden Sie das am nächsten verwandte menschliche Protein.
- Wiederholen Sie die Suche, indem Sie die BLOSUM250 Scoring Matrix verwenden.

12. Haben die folgenden Paare von Bäumen identische Topologie?

a)



b)



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 | |

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