Übungsblatt 3

**3.1**

1a) p-Distanz (Hamming-Distanz)

P=D/L

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 |
| 1 | 0 | 0.2 | 1 | 0 |
| 2 | 0.2 | 0 | 9.8 | 0.2 |
| 3 | 1 | 0.8 | 0 | 1 |
| 4 | 0 | 0.2 | 1 | 0 |

b) Poisson-korrigierte Distanz

dp = -ln(1-p) ; für kleine p zu ignorieren, für große p wichtig

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 |
| 1 | 0 | 0.223 | - | 0 |
| 2 | 0.223 | 0 | 1.609 | 0.223 |
| 3 | - | 1.609 | 0 | - |
| 4 | 0 | 0.223 | - | 0 |

für p=1 ist die Poisson-korrigierte Distanz undefiniert

c) Jukes-Cantor Distanz

dJC = -(3/4) ln(1-(4/3)p) ; nicht gültig für hohe p

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 |
| 1 | 0 | 0.233 | - | 0 |
| 2 | 0.233 | 0 | - | 0.233 |
| 3 | - | - | 0 | - |
| 4 | 0 | 0.233 | - | 0 |

Hier: nicht gültig für p=0.8 und p=1

Die p-Distanz – die geschätzte Wahrscheinlichkeit, dass eine Position mutiert ist, unterschätzt die wahre Distanz zweier Sequenzen, da nicht berücksichtigt wird, dass wiederholte Mutationen an der gleichen Stelle auftreten können. Durch die Poisson-Korrektur wird diese Möglichkeit berücksichtigt, was sich dann in größeren Distanzwerten niederschlägt. Für kleine p-Distanzen ist die Korrektur recht klein, hier zum Beispiel erhöht sich die Distanz von 0.2 auf 0.223. Der Wert für die p-Distanz mit Poisson-Korrektur steigt durch den logarithmischen Ausdruck mit größer werdendem p stark an. So erhöht sich die p-Distanz in dem hier gegebenen Beispiel durch die Poisson-Korrektur von 0.8 auf über 1.6.

Die Jukes-Cantor-Distanz stellt auch ein korrigiertes Abstandsmaß dar. In diesem Modell wird die Annahme gemacht, dass die 4 Nukleotide A, T, C und G gleich häufig in der DNA-Sequenz auftreten und dass sich jedes Nukleotid mit der gleichen Wahrscheinlichkeit in ein anderes Nukleotid umwandelt. Die Limitation des Jukes-Cantor-Modells ist, dass es nur für ähnliche Sequenzen anwendbar ist: für ist die evolutionäre Distanz nicht definiert.

**3.1.2**

I‘=I/100

I‘ Übereinstimmung

I Prozentuale Übereinstimmung

Beweise, dass die Poisson-korrigierte Distanz zwischen zwei Sequenzen

dp =-ln(I‘) ist

Definition der Poisson-korrigierten Distanz

,

mit

D 🡪 Anzahl an mismatches

M 🡪 Anzahl an matches

mit

**3.2.1**

Bei dieser Matrix handelt es sich um eine valide Matrix. Die Identität des Ununterscheidbaren d(a,a)=0 ist gegeben und die Matrix ist symmetrisch.

**3.2.2**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| A | 0 | 5 | 9 | 9 | 8 |
| B | 5 | 0 | 10 | 10 | 9 |
| C | 9 | 10 | 0 | 8 | 7 |
| D | 9 | 10 | 8 | 0 | 3 |
| E | 8 | 9 | 7 | 3 | 0 |

D

E

W

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A | B | C | W |
| A | 0 | 5 | 9 | 8.5 |
| B | 5 | 0 | 10 | 9.5 |
| C | 9 | 10 | 0 | 7.5 |

A

B

X

|  |  |  |  |
| --- | --- | --- | --- |
|  | C | W | X |
| C | 0 | 7.5 | 9.5 |
| W | 7.5 | 0 | 9 |

C

D

E

W

Y

|  |  |  |
| --- | --- | --- |
|  | X | Y |
| X | 0 |  |

C

D

E

W

Y

A

B

X

Z

**3.3**

Diagram

Description automatically generated with medium confidence

Chart, radar chart

Description automatically generated

UPGMA (=unweighted pair group method with arithmetic mean) ist ein iterativer Algorithmus, der zur Erstellung phylogenetischer Bäume genutzt werden kann. In dem hier gegebenen Beispiel erlaubt der Algorithmus die Konstruktion eines ultrametrischen Baumes, der die Richtung evolutionärer Verhältnisse angibt. Bei der Konstruktion eines ultrametrischen Baumes wird angenommen, dass die Rate der Mutationen konstant ist; man spricht auch von einer Molekularen Uhr. Die Vereinfachung, eine räumlich und zeitlich konstante Mutationsrate anzunehmen, ist nützlich, trifft aber meistens nicht zu, beispielsweise weil die dritte Codon Position häufiger mutiert ist oder da Umwelteinflüsse temporär einen größeren Selektionsdruck auf bestimmte Gene erzeugen.

Der Neighbour-Joining Algorithmus hingegen berücksichtigt, dass die Mutationsrate und somit die Evolutionsgeschwindigkeit nicht konstant ist. In dem gegebenen Beispiel wurde mit dem Algorithmus ein wurzelfreier Baum erzeugt, der keine Aussagen über die Richtung evolutionärer Verhältnisse zulässt. (Es kann jedoch auch ein Baum erstellt werden, aus dem der Urvorfahr hervorgeht).

3.3.1

Unter den Spezies befinden sich 2 Nicht-Primaten: die Maus und das Rind. Im gewurzelten UPGMA-Baum ist die Maus klar von den Primaten abgegrenzt, da sie sich nicht in Sub-Clustern mit Primaten befindet. Das Rind hingegen ist im UPGMA-Baum in Sub-Clustern mit Primaten lokalisiert. Im Gegensatz zum UPGMA-Baum ist im ungewurzelten NJ-Baum erkennbar, dass Lemuren und Tarsier clustern, Rinder und Mäuse jedoch nicht deutlich mit Primaten-Arten clustern. Die zwei großen Cluster um Gorilla etc. und die Macaques – im UPGMA-Baum erkennbar – sind auch im NJ-Baum zu finden. Die Einordnung des Rindes in ein Primaten-Cluster im UPGMA-Baum könnte auf die Annahme einer konstanten Mutationsrate (eine Limitation der Molekulare Uhr) zurückzuführen sein. Alternativ könnte es auch sein, dass es sich bei den zu Grunde liegenden DNA-Sequenzen um solche handelt, die durch konvergente Evolution tatsächlich eine große Ähnlichkeit zwischen Rindern und Primaten-Arten aufweisen.

3.3.2

Während im UPGMA-Baum Lemuren mit Rindern und Tarsiern ein Cluster bilden, wobei die evolutionäre Distanz zwischen Lemuren und Rindern kleiner ist als die Distanz beider zu Tarsiern, geht aus dem ungewurzelten NJ-Baum hervor, dass Lemuren den Tarsiern evolutionär nahe stehen, Rinder jedoch etwas weiter entfernt sind, sodass unklar ist, ob von einem Cluster gesprochen werden kann. Im NJ-Baum ist die Primaten-Art, welche dem Tarsier evolutionär am zweitnächsten steht, der Squirrel Monkey.

3.3.3

Aus beiden phylogenetischen Bäumen geht hervor, dass der Mensch mit dem Schimpansen ein Cluster bildet.