1	2	3	$\sum$ (7)

Bioinformatics I
WS 15/16
Tutor: Alexander Seitz

Jonas Ditz
& Benjamin Schroeder

Blatt 11

(Abgabe am 1. Februar 2016)

1 Theoretical Assignment - *Coverage statistics*

1.1

1.2

1.3

2 Theoretical Assignment - *Application of the arrival statistic for unitigs*

3 Theoretical Assignment - *On distances*

3.1 title

Consider a tree T constructed with D and the four nodes $i, j, k, l \in T$. Since D is ultrametric, the following four inequalities have to hold:

$$d(i,j) \leq max\left\{d(i,k), d(j,k)\right\}$$
$$d(i,j) \leq max\left\{d(i,l), d(j,l)\right\}$$
$$d(k,l) \leq max\left\{d(i,k), d(i,l)\right\}$$
$$d(k,l) \leq max\left\{d(j,k), d(j,l)\right\}$$

W.l.o.g we can assume that $d(i,k) = d(j,k) = d(i,l) = d(j,l)$. With that assumption we can rewrite the inequalities as

$$d(i,j) + d(i,j) \leq d(i,k) + d(j,l)$$
$$d(k,l) + d(k,l) \leq d(i,l) + d(j,k)$$

And also the following inequality is valid:

$$d(i,j) + d(k,l) \leq max\left\{d(i,k) + d(j,l), d(i,l) + d(j,k)\right\}$$

As one can see, that is the Four-Point-Condition (4PC) and a metric fulfill 4PC if and only if it is additive. So D is a tree metric $\square$

To show the back direction one consider the four elements $A, B, C, D \in X$ of a taxa X and the following distance matrix:

$$D = \begin{array}{cc} & \begin{array}{ccc} B & C & D \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 7 & 6 & 5 \\ & 3 & 6 \\ & & 5 \end{bmatrix} \end{array}$$

One can see in the script of this lecture that D is a tree metric. But the Three-Point-Condition (3PC) is not fulfilled as can be easily shown. To fulfill 3PC the following inequality has to be valid:

$$d(A, B) \leq max\,\{d(A, C), d(B, C)\}$$

but

$$7 \nleq max\,\{6, 3\}$$

So D does not fulfill 3PC and thus D is not ultra metric. □

3.2

References