**Praktika z bioinformatiky**

Fylogenetika

Jméno:

Příjmení:

1. Vysvětlete pojem OTU
2. Jaký je rozdíl mezi fylogramem a dendrogramem?
3. Popište rozdíl mezi znakovými a distančními metodami pro tvorbu fylogramu.
4. Vysvětlete pojem bifurkovaný strom.
5. Popište alespoň 4 distanční metriky.
6. Popište alespoň 4 shlukovací metody pro tvorbu stromů.
7. Popište princip shlukování metodou UPGMA.
8. Sestrojte strom metodou UPGMA pro následující sekvence. Pro výpočet distanční matice použijte p-distanci.

1: ACTGCTTGCA

2: ACACCTAGCA

3: TGTGCGTGCC

4: ACGGTCAGTG

1. Vytvořený strom zapište v NEWICK formátu s délkami větví.
2. Nainstalujte si knihovnu ‘ape’ a ‘msa’ (install.packages(‘ape’), BiocManager::install("msa")). Proč se instalují rozdílnými příkazy?
3. Z předchozího cvičení si stáhněte soubory mitochondriálních genů. Vyberte si 5 organismů a extrahujte z nich gen COX2. Výsledný soubor uložte jako samostatnou fastu.
4. Sekvence přeložte do proteinu a také uložte. Pozor na zvolení správného genetického kódu.
5. DNA i AA zarovnejte pomací nástroje ClustalOmega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>)
6. Zarovnání načtěte v R (readDNAMultipleAlignment()). Pomocí funkce stringDist() spočítejte distanční matici.
7. Strom vykreslete pomocí funkcí hclust() a plot(). Nezapomeňte na správné pojmenování.
8. Váš strom zapište na disk ve formátu NEWICK. (write.tree(as.phylo(), file=” ”)).
9. Sekvence zarovnejte pomocí knihovny msa a opakujte krok 15-16. Porovnejte eůzné metody zarovnání a metody konstrukce stromů.