Assignment 4 / Biometry and bioinformatics I / 2013

Submission deadline 31.12.2013

Topic: Mammalian classification by sequence information

Construct a sequence string from four genes (concatenate the sequences), align them and construct a UPGMA-tree (see the note in assignment 3; you may also do some other phylogenetic clustering which you learn soon in BB_II).

Use ~15 mammalian species. You may already have a set from assignment 1. You can use that, or some other set of your choice.

Consult the "official tree of mammals" from here: http://tolweb.org/tree/
and write a report which includes an interpretation of your results in the face of this information (how well does a set of four genes reflect the "true" relationships of mammals).

The genes

gene_1.txt is the gene from assignment 1
gene_2.txt is the gene from the tutorial example, from the same species

Genes 1 and 2 are from the nuclear genome.

In addition, use two genes from the mitochondrial genome (mt-genome) which is a "miniature genome" , and often used as a marker for various purposes.

gene_3.txt is from mt-genome, from the same species as 1 and 2.
gene_4: take it also from the mitochondrial genome, which is given as mt_genomes_aligned.txt (a set of mammals). Take the gene 16S ribosomal RNA. Use the same species as above (using the same species = you have the same kind of starting condition in blasting additional data). You find the coordinates for this gene from the table given below.

This assignment includes several kind of practical problems to be solved – intentionally => learning to work with sequences!

	Nucleotides in AB499817, the first sequence in datafile	Nucleotides taking into account gaps in aligned file
tRNA-Phe	1-69	1-80
12S ribosomal RNA	70-1023	81-1090
tRNA-Val	1024-1090	1091-1161
16S ribosomal RNA	1091-2670	1162-2840
tRNA-Leu	2671-2745	2841-2917
gene ND1	2748-3704	2919-3882
tRNA-Ile	3704-3722	3882-3901
tRNA-Gln	3769-3843	3948-4025
tRNA-Met	3845-3914	4028-4097
gene ND2	3915-4958	4098-5143
tRNA-Trp	4957-5024	5142-5215
tRNA-Ala	5038-5106	5232-5301
tRNA-Asn	5108-5179	5310-5386
tRNA-Cys	5213-5280	5419-5495
tRNA-Tyr	5281-5348	5496-5572
gene COI	5350-6894	5574-7140
tRNA-Ser	6892-6962	7132-7216
tRNA-Asp	6967-7034	7222-7292
gene COII	7035-7718	7293-7977
tRNA-Lys	7736-7802	7995-8066
gene ATPase subunit 8	7804-8007	8068-8276
gene ATPase subunit 6	7965-8645	8234-8914
gene COIII	8645-9428	8914-9697
tRNA-Gly	9429-9496	9698-9770
gene ND3	9497-9843	9771-10117
tRNA-Arg	9843-9911	10117-10187
gene ND4L	9914-10210	10191-10487
gene ND4	10204-11581	10481-11858
tRNA-His	11580-11650	11857-11930
tRNA-Ser	11651-11710	11931-11995
tRNA-Leu	11711-11780	11996-12067
gene ND5	11781-13601	12068-13895
gene ND6	13585-14112	13879-14406
tRNA-Gln	14111-14181	14405-14476
gene cytB	14186-15325	14482-15625
tRNA-Thr	15326-15395	15626-15703
tRNA-Pro	15395-15460	15703-15772
D-loop	15461-16741	15773-18424