Mini projekt 2. Kateri geni nam lahko pomagajo razločiti med ribo in sesalci?

Ivan Antešić (63130003)

December 27, 2019

1 Odgovori na vprašanja

Vseh 20 genov smo uredili po sposobnosti razločitve med dvema skupinama (ribo Danio rerio in preostalimi šestimi sesalci) od najboljšega do najslabšega. To smo dosegli z mero, ki izračuna razdaljo med zaporedjem ribe in enotnim zaporedjem (angl. consensus) vseh sesalcev. Razdalja med zaporedjema je v navodilih naloge opisana Jukes-Cantor popravljena razdalja. Ker so zaporedja za različne vrste različnih dolžin smo hevristično določili, da je dolžina enotnega zaporedja enaka najkrajšemu zaporedju vrste in ignorirali preostanek drugih zaporedij. Rezultati razvrščanja so vidni v tabeli 1. Nato smo za prvo in zadnje uvrščena gena pridobili filogenetsko drevo (sliki 1 in 2)

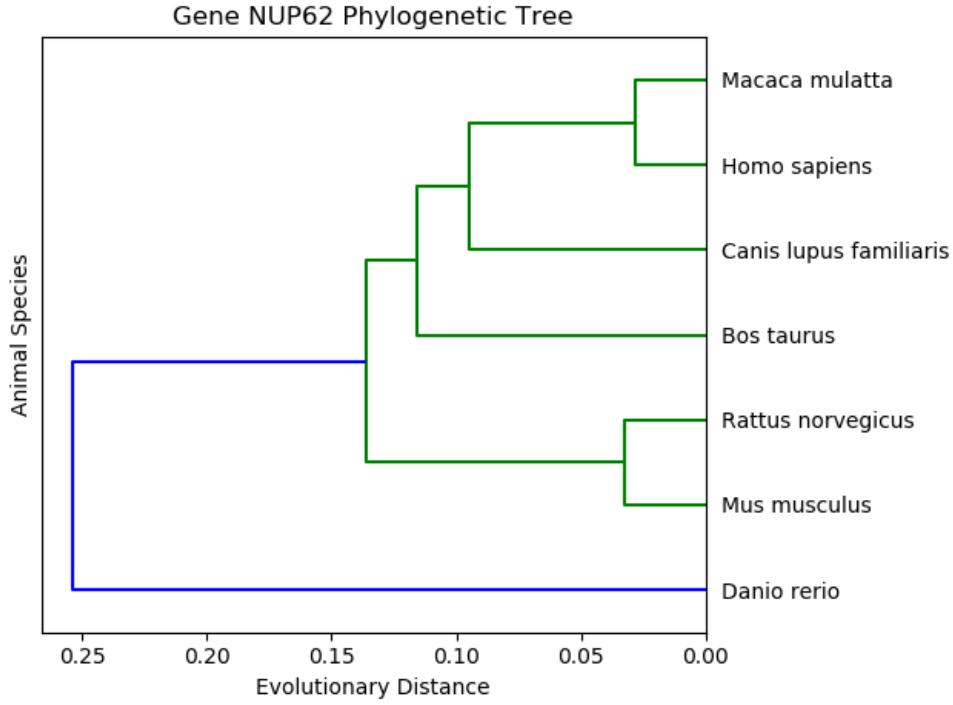

Figure 1: Filogenetsko drevo za najbolj diskriminanten gen glede na našo razvrstitev.

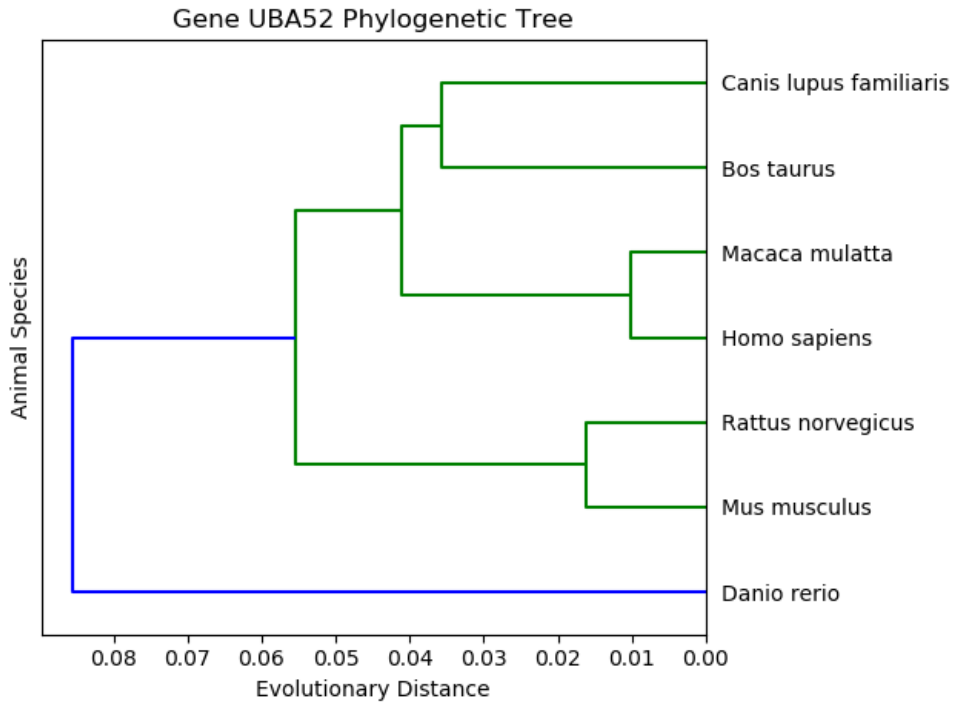

Figure 2: Filogenetsko drevo za najmanj diskriminanten gen glede na našo razvrstitev.

2 Rezultati

Morda bi bilo za določitev enotnega zaporedja bolje uporabiti večkratno poravnavo (angl. multiple alingment) vendar bi v tem primeru imeli večjo kompleksnost algoritma. Če primerjamo razvrstitev genov z rezultati mere, ki izračuna povprečno razdaljo med ribo in vsemi člani skupine sesalcev lahko opazimo, da se razvrstitvi ne razlikujeta bistveno (tabela 1). Za razvrstitev glede na povprečno razdaljo je bilo potrebnih popvrečno 700 sekund in 120 izračunov razdalj (za vsak gen 6), medtem ko je bilo za razvrstitev z enotnim zaporedjem potrebnih le 120 sekund in 20 izračunov razdalj (za vsak gen 1).

Izračun razdalje med dolgima zaporedjema se je zaradi iskanja globalne poravnave v pythonu izkazal za časovno zahtevno operacijo. Zato smo uporabili razvrstitev z enotnim zaporedjem. Morda bi bilo boljše za vsak gen določiti filogenetsko drevo (angl. phylogenetic tree) in za mero uporabiti razmerje razdalje pri kateri združimo ribo s sesalci in razdaljo pri kateri so združeni vsi sesalci - torej razvrstitev glede na velikost razmaka med sesalci in ribo in ne glede na razdaljo združitve. V tem primeru bi se gen EXT1 (slika 3) izkazal za zelo diskriminativnega, čeprav ga naša razvrstitev uvršča na 14. mesto. Vendar bi bila takšna reštev časovno potratna saj je bilo samo za izračun matrike razdalj med vrstami za gen EXT1 potrebnih 452 sekund.

mesto	razvrstitev z enotnim zaporedjem	razvrstitev s povprečno razdaljo
1	NUP62	NUP62
2	DNASE1	S100A4
3	S100A4	DNASE1
4	SH3KBP1	MRPL21
5	MNS1	SH3KBP1
6	MRPL21	MNS1
7	INA	INA
8	LGI1	LGI1
9	BMP4	BMP4
10	RPS6KA3	DLX5
11	GBX2	GBX2
12	DLX5	RPS6KA3
13	VAMP2	VAMP2
14	EXT1	EXT1
15	RPL35	RPL35
16	RAC2	RAC2
17	RPL7A	RPL18A
18	RPL18A	RPL7A
19	RPL39	RPL39
20	UBA52	UBA52

Table 1: Primerjava razvrestitve z enotnim zaporedjem in razvrstitve s povprečno razdaljo.

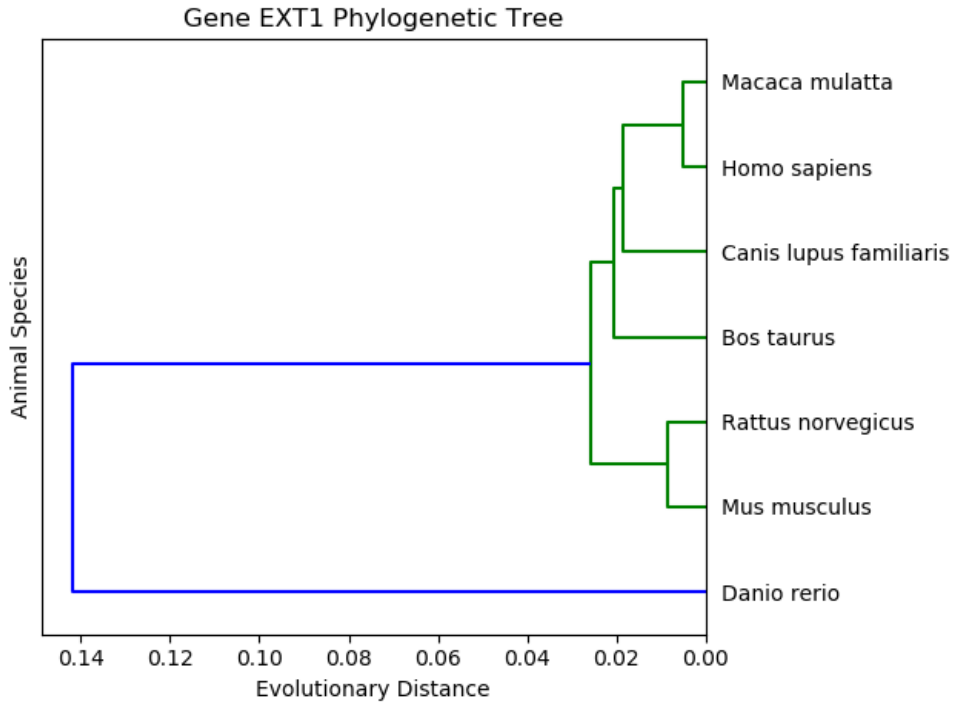

Figure 3: V primerjavi s drevesom na sliki 1 vidimo, da gen dobro diskriminira med ribo in sesalci, kljub temu, da je zadnja združitev pri manjši razdalji.