

Matas Bobin

Bioinformatikos lab 1 ataskaita

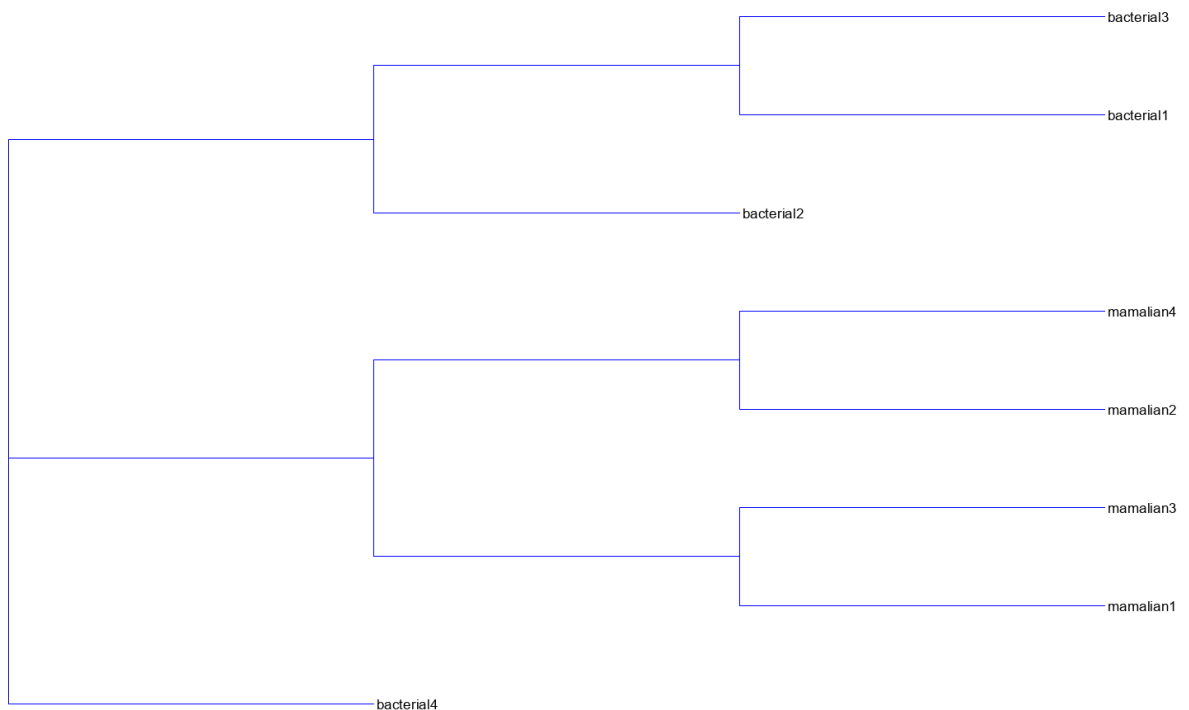
1. Atstumo funkcija

Atstumo funkcija buvo skaičiuojama naudojant Euklido atstumą tarp kiekvienos sekos kodonų ir dikodonų dažnių vektorių. Ši funkcija veikia taip:

- a) Daznių skaičiavimas: Kiekvienai sekai buvo apskaičiuoti visi galimi kodonų ir dikodonų dažniai.
- b) Vektorizavimas: Kodonų ir dikodonų dažniai buvo paversti vektoriais, kuriuose kiekvienas elementas atitinka tam tikro kodono ar dikodono dažnį.
- c) Atstumo skaičiavimas: Naudojant Euklido metriką, buvo apskaičiuotas atstumas tarp dažnių vektorių, kuris parodo skirtumą tarp sekų pagal jų kodonų ir dikodonų naudojimo modelius.

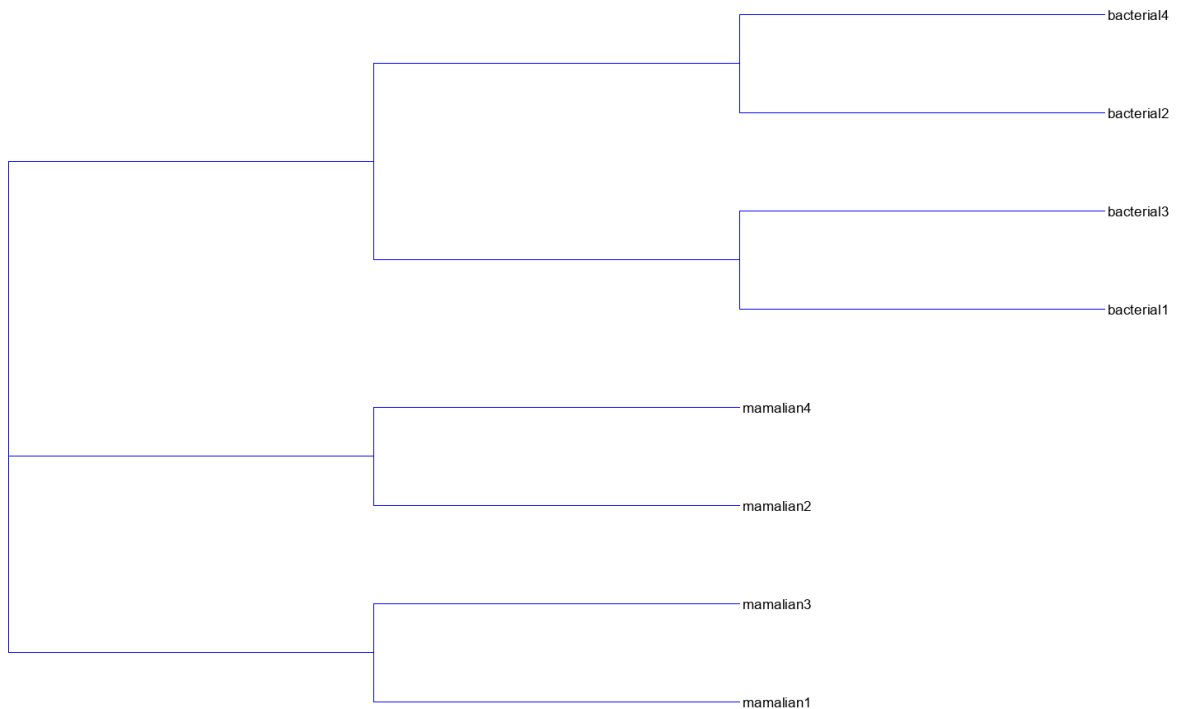
2. Medžiai:

a) Kodonų medis:



Kodonų medis parodo aiškų pasidalinimą tarp bakterijų ir žinduolių virusų. Žinduolių virusai klasterizuojasi vienoje šakoje, o bakterijų virusai – kitoje.

b) Dikodonų medis:



Dikodonų medis taip pat rodo skirtumą tarp bakterijų ir žinduolių virusų, tačiau klasterizacija yra šiek tiek kitokia. Pavyzdžiui, vienas iš bakterijų virusų (bacterial4) šiek tiek labiau artimas žinduolių virusams pagal dikodonų dažnį, tačiau vis tiek aiškiai priklauso bakterijų grupei.

3. Kodonų ir dikodonų dažniai tarp žinduolių ir bakterijų virusų skiriasi gana ryškiai. Šie skirtumai pastebimi tiek kodonų, tiek dikodonų dažnių atstumo matricose:

a) Kodonų atstumo matrica:

mamalian1	0.000	0.247	0.190	0.310	0.264	0.230	0.307	0.214
mamalian2	0.247	0.000	0.197	0.183	0.199	0.182	0.250	0.172
mamalian3	0.190	0.197	0.000	0.263	0.246	0.222	0.303	0.186
mamalian4	0.310	0.183	0.263	0.000	0.248	0.259	0.308	0.231
bacterial1	0.264	0.199	0.246	0.248	0.000	0.131	0.144	0.136
bacterial2	0.230	0.182	0.222	0.259	0.131	0.000	0.205	0.105
bacterial3	0.307	0.250	0.303	0.308	0.144	0.205	0.000	0.215
bacterial4	0.214	0.172	0.186	0.231	0.136	0.105	0.215	0.000

b) Dikodonų atstumo matrica:

mamalian1	0.000	0.172	0.165	0.185	0.198	0.175	0.201	0.168
mamalian2	0.172	0.000	0.169	0.168	0.186	0.172	0.202	0.162
mamalian3	0.165	0.169	0.000	0.190	0.193	0.175	0.214	0.163
mamalian4	0.185	0.168	0.190	0.000	0.189	0.179	0.205	0.167
bacterial1	0.198	0.186	0.193	0.189	0.000	0.174	0.181	0.159
bacterial2	0.175	0.172	0.175	0.179	0.174	0.000	0.194	0.137
bacterial3	0.201	0.202	0.214	0.205	0.181	0.194	0.000	0.181
bacterial4	0.168	0.162	0.163	0.167	0.159	0.137	0.181	0.000

Iš abiejų matricų aiškiai matyti, kad žinduolių ir bakterijų virusai formuoja atskiras klasterių grupes.

Žinduolių virusai (mamalian1, mamalian2, mamalian3, mamalian4) formuoja atskirą grupę tiek kodonų, tiek dikodonų matricose. Jie turi panašius kodonų ir dikodonų dažnius tarpusavyje, bet didesnius skirtumus nuo bakterijų virusų.

Bakterijų virusai (bacterial1, bacterial2, bacterial3, bacterial4) taip pat formuoja atskirą grupę, kurioje tarpusavio atstumai yra mažesni, rodo panašius kodonų ir dikodonų naudojimo modelius.

Bakterijų virusas bacterial4 išsiskiria tiek kodonų, tiek dikodonų matricose, nes jo atstumai iki kitų bakterijų virusų (pvz., bacterial1 ir bacterial2) yra šiek tiek didesni nei kitų bakterijų virusų tarpusavio atstumai. Kodonų atveju, bacterial4 turi didesnę atstumą (pvz., 0.214 iki mamalian1), tačiau dikodonų atstumo matricoje jis yra artimesnis kai kuriems žinduolių virusams (pvz., 0.168 iki mamalian1). Tai rodo, kad bacterial4 turi tam tikrų ypatumų, dėl kurių jis šiek tiek artimesnis žinduolių virusams pagal dikodonų naudojimo modelį.

4. Labiausiai varijuojantys kodonai yra:

- Glutamic acid
- Arginine
- Lysine

Labiausiai varijuojantys dikodonai yra:

- Glutamic acid-Leucine
- Isoleucine-Glutamic acid
- Threonine-Glutamic acid