

Lab 1 - Ataskaita

Domas Kalinauskas

Atstumo funkcija

Atstumo funkcijai parinkau euklido atstumą, t.y.

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

(kur n yra 7 - kodono, ir 400 - dikodono atvėju)

šia operacija atitinka `scipy` bibliotekos `pdist` funkcija. Ji gražina masyvą vektorių, tai poto naudojant `squareform` paverčiam į atstumo matricos formą.

Kodonų ir Dikodonų medžiai

Gauti kodonų ir dikodonų medžiai pagal klusterizavimąsi atrodo identiški (tačiau atstumų matricių vertės ne vienodos!). Žr. Figure 1 bei Figure 2 dokumento apačioje.

Dažniai, klasterizavimas

Ar skiriasi kodonų ir dikodonų dažnis tarp žinduolių ir bakterijų virusų?

Taip, kodonų ir dikodonų dažniai skiriasi. Tai matosi iš to, kad atstumai tarp bakterijų ir žinduolių virusų yra didesni, negu jų grupėse (t.y. bakterijos vs bakterijos atstumas bendru atvėju mažesnis negu bakterija vs virusas, ir atsvirkščiai).

Kaip klasterizuoja virusai?

Bendru atvėju, virusai klasterizuoja pagal jiems priklausančią kategoriją (t.y. bakterijų su bakterijų, žinduolių su žinduolių).

Gal kažkuris virusas labai išsiskyrė?

Išsiskyrė virusas U18337.1 (`mamalian3.fasta`), kadangi jis klasterizavosi arčiau bakterinių, vietoj žinduolinių (buvo arčiau 3/4 bakterinių negu su likusiais žinduolių).

Kokie kodonai/dikodonai labiausiai varijuoja?

Didžiausia variacija buvo tarp herpesvirus (`mamalian4.fasta`) ir U183371 (`mamalian3.fasta`). Kodonų matricioje atstumas buvo 0.124, dikodonų matricioje atstumas buvo 0.048.

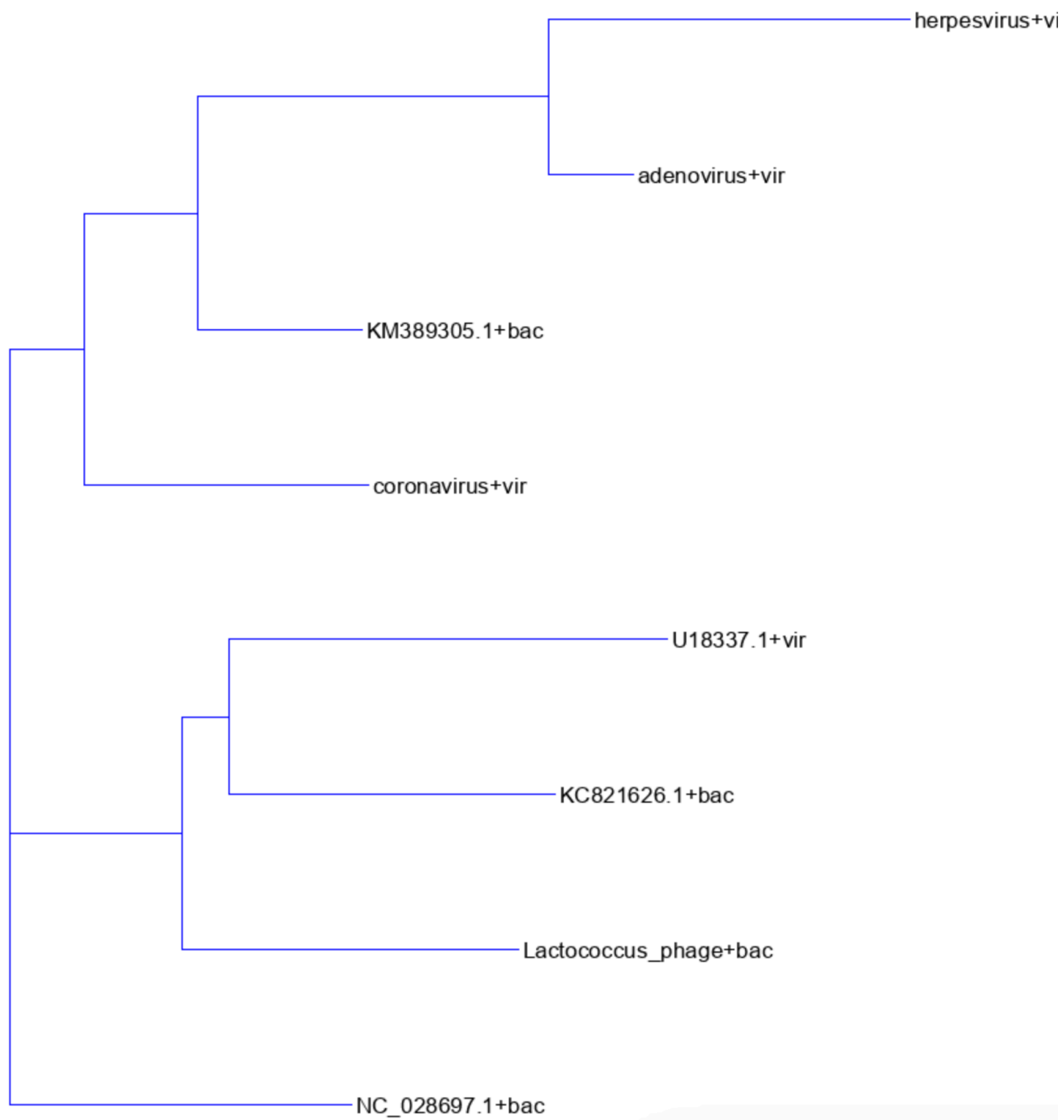


Figure 1: Kodonai

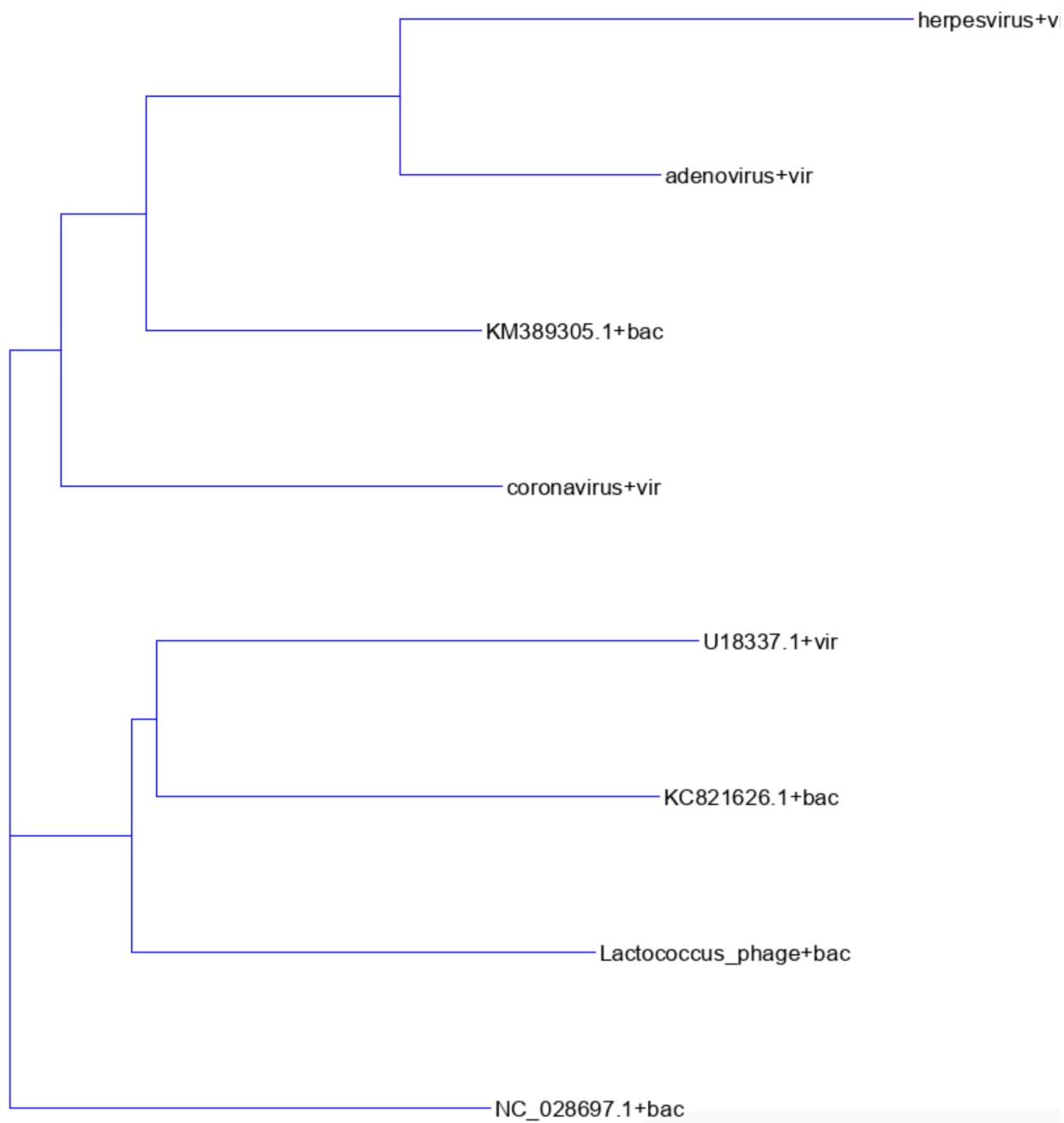


Figure 2: Dikodonai