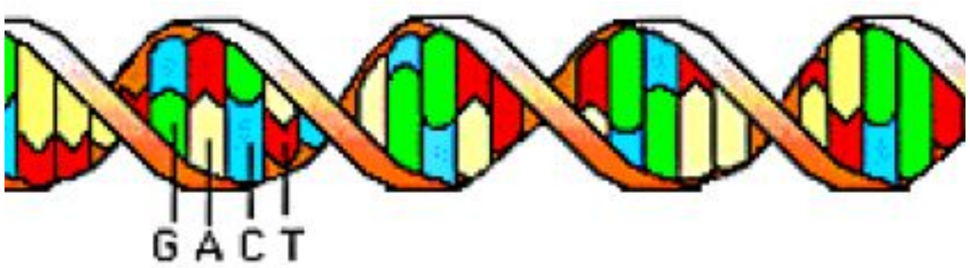
Bazės

4 rūšys molekulių, dvi grupės

Adeninas(A), Guaninas(G) ← Purinai

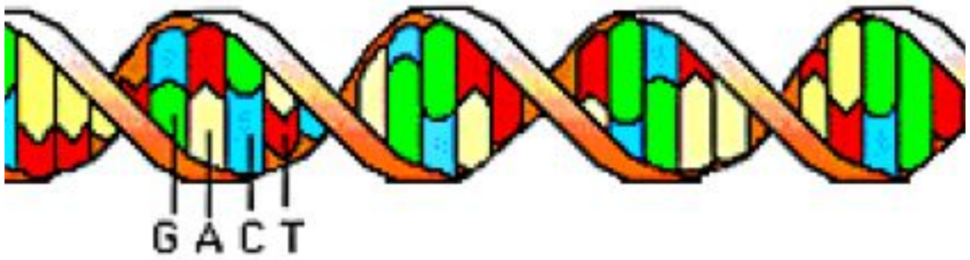
Citozinas(C), Timinas(T) ← Pirimidinai



Bazės

A visada stovi prieš T

G visada stovi prieš C



Bazės

A visada stovi prieš T

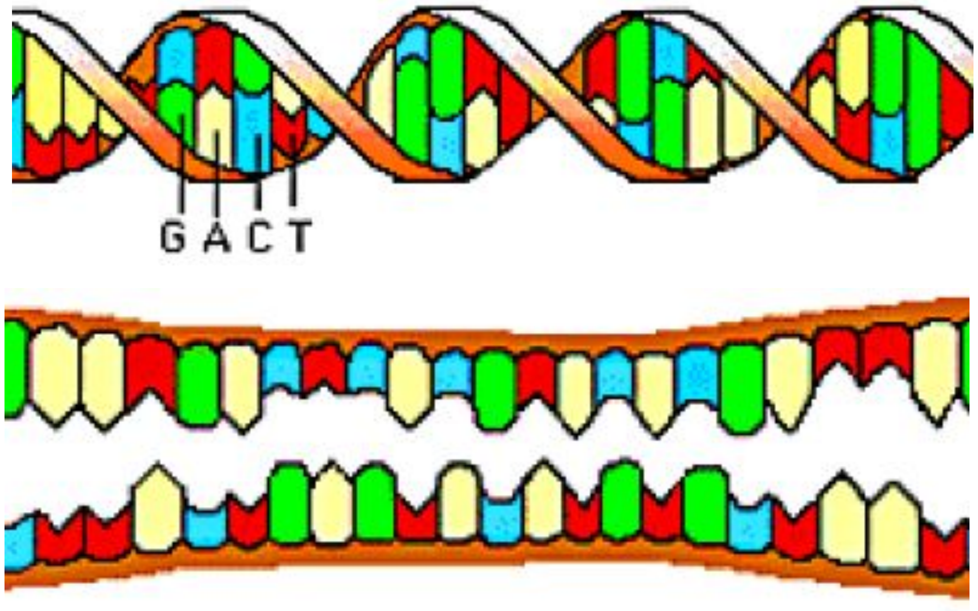
G visada stovi prieš C

Seka: AGCGCT

Komplementari seka: TCGCGA

Bazės

- Tam kad apibūdinti DNR seką užtenka išvadinti DNR bazes vienoje grandinėje.



Mutacijos

- Mutacijos DNR sekoje atsiranda atsitiktinai. Žmogus turi apie ~40 mutacijų atsitiktinai atsiradusių ir nerastų nei viename tėve.

Mutacijos

- Mutacijos DNR sekoje atsiranda atsitiktinai.

Bazių pakeitimas

- Dažniausia mutacijų forma
- Bazė yra pakeičiama kita baze

$$\begin{array}{cccccc} A & A & T & C & G & C \\ \downarrow & & & \downarrow & & \\ G & A & T & G & G & C \end{array}$$

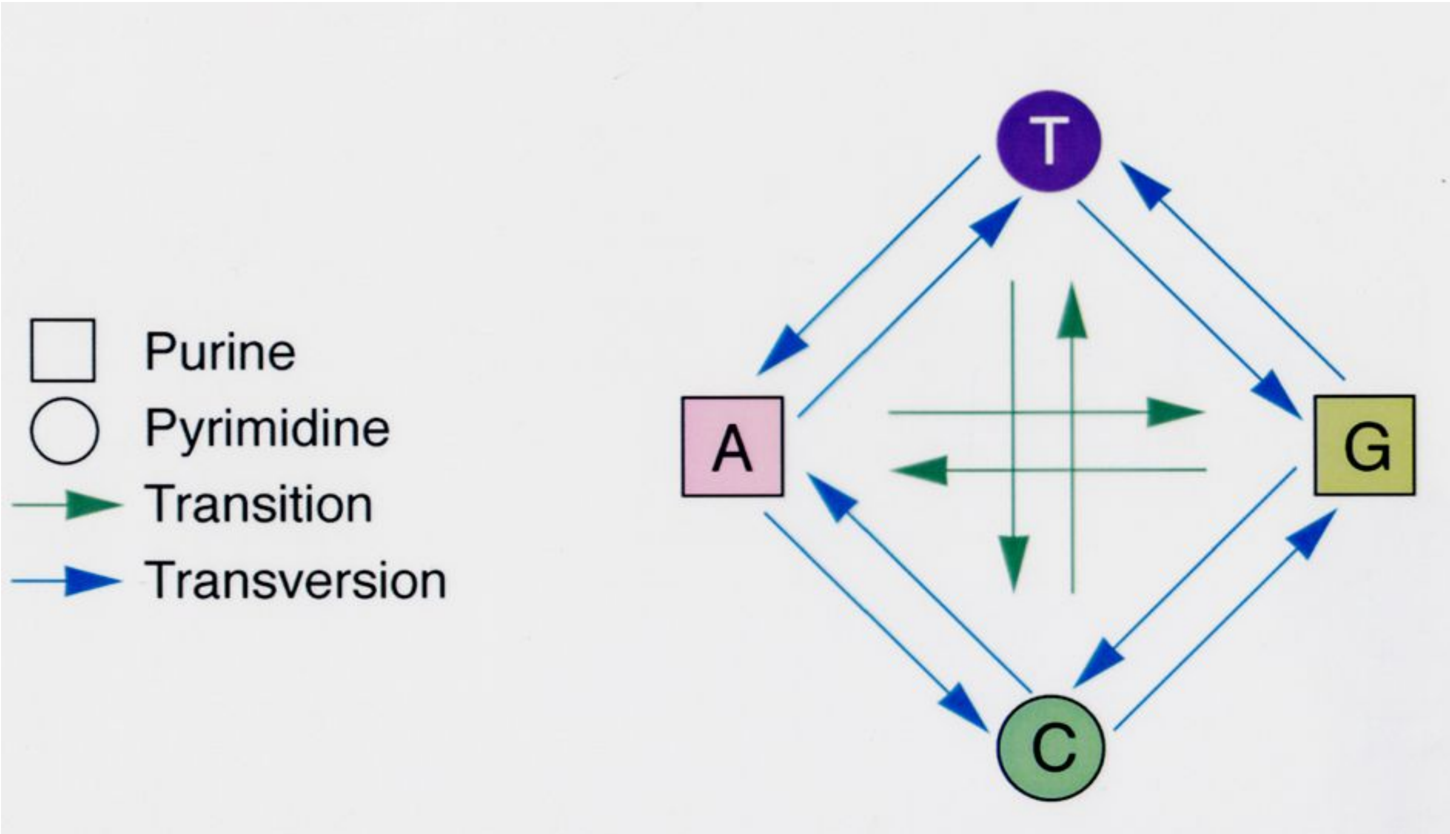
Bazių pakeitimas

- Tranzicija: Pur į Pur, Pyr į Pyr
- Transversija: Pur į Pyr, Pyr į Pur

$$\begin{array}{cccccc} A & A & T & C & G & C \\ \downarrow & & & \downarrow & & \\ G & A & T & G & G & C \end{array}$$

Tranzicija Trasnsversija

Bazių pakeitimas



Esminis klausimas

- Kaip nuspręsti kiek mutacijų skiria sekas?

Pavyzdys

S0 : Bendras protėvis

S1 : Palikuonis sekos S0

S2 : Palikuonis sekos S1

S0 : ATGTCGCCTGATAATGCC…

S2 : ATGCCGCGTGATAATGCC…

Pavyzdys

S0 : Bendras protėvis

S1 : Palikuonis sekos S0

S2 : Palikuonis sekos S1

S0 : ATGTCGCCTGATAATGCC…

S2 : ATGCCGCGTGATAATGCC…

Stebimų mutacijų skaičius: 2

Pavyzdys

S0 : Bendras protėvis

S1 : Palikuonis sekos S0

S2 : Palikuonis sekos S1

S0 : ATGTCGCCTGATAATGCC…

S1 : ATGCCGCTTGACAATGCC…

S2 : ATGCCGCGTGATAATGCC…

Faktinis mutacijų skaičius: 5

Pavyzdys

S0 : Bendras protėvis

S1 : Palikuonis sekos S0

S2 : Palikuonis sekos S1

S0 : ATGTCGCCTGATAATGCC…
S1 : ATGCCGCTTGACAATGCC…
S2 : ATGCCGCGTGATAATGCC…

Faktinis mutacijų skaičius: 5

(dalis yra *paslėptos mutacijos*)

Ko mes norime?

- Palyginti pradinę ir galutinę DNR seką?
- Sukurti matematinį modelį rasti visas įvykusias mutacijas įtraukiant ir paslėptas mutacijas.

Reality...

- Beveik niekada neturime protėvio sekos. Tik kartais turime keletą susijusių organizmų iš praeities.
- Dažniausiai turime sekas tik dabar gyvenančių palikuonių, bet niekada neturime informacijos apie dabar neegzistuojančių organizmų sekas.

Realybė...

- Kai mes lyginame dvi sekas ir bandome atkurti mutacijas, kurioms vykstant jos atsirado, mes nežinome ***paskutinio bendro protėvio*** sekos.

Ortologinės sekos

- Duotai DNR sekai mes galime rasti panašias sekas DNR duombazėse.
- Jei genas randamas viename organizme, mes galme greitai identifikuoti panašių genų sekas susijusiuose organizmuose.

Ortologinės sekos

- Jei genas turi panašią funkciją, mes galime pagrįstai teikgti kad jos kilo iš bendro protėvio ir yra *ortologinė.*

Sąlyginė tikimybė

Apibrėžimas

Esant dviems galimiems įvykiams E ir F, sąlyginė tikimybė kad įvyks įvykis E įvykus įvykiui F apibrėžiama taip:

$$P(F \mid E) = \frac{P(F \cap E)}{P(E)}$$

Pavyzdys

Tarkim yra 40-bazių ilgio protėvio seka S_0 ir palikuonio seka S_1

$$S_0 : ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT$$

$$S_1 : ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC$$

Example

Suskaičiuojam įvykusius bazių pakeitimus:

S_0 : *ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT*

S_1 : *ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC*

$S_1 \setminus S_0$	A	G	C	T
A	7	0	1	1
G	1	9	2	0
C	0	2	7	2
T	1	0	1	6

Example

Galime įvertinti $P(S_1 = i \mid S_0 = j)$

$S_0 : ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT$

$S_1 : ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC$

$S_1 \setminus S_0$	A	G	C	T
A	7	0	1	1
G	1	9	2	0
C	0	2	7	2
T	1	0	1	6

Example

Galime įvertinti $P(S_1 = i \mid S_0 = j)$

$S_0: ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT$

$S_1: ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC$

$S_1 \setminus S_0$	A
A	7
G	1
C	0
T	1

$$P(S_1 = A \mid S_0 = A) = \frac{7}{9}$$

$$P(S_1 = G \mid S_0 = A) = \frac{1}{9}$$

$$P(S_1 = C \mid S_0 = A) = 0$$

$$P(S_1 = T \mid S_0 = A) = \frac{1}{9}$$

Example

Q1: Kokia suma yra 16-os langelių? Kodėl?

S_0 : $ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT$

S_1 : $ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC$

$S_1 \setminus S_0$	A	G	C	T
A	7	0	1	1
G	1	9	2	0
C	0	2	7	2
T	1	0	1	6

Example

Kam lygios eilučių sumos?

$$S_0 : ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT$$

$$S_1 : ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC$$

$S_1 \setminus S_0$	A	G	C	T
A	7	0	1	1
G	1	9	2	0
C	0	2	7	2
T	1	0	1	6

Example

Galime sudaryti sąlyginės tikimybės matricą $P(S_1 = i \mid S_0 = j)$

$S_1 \setminus S_0$	A	G	C	T
A	$\frac{7}{9}$	0	$\frac{1}{11}$	$\frac{1}{9}$
G	$\frac{1}{9}$	$\frac{9}{11}$	$\frac{2}{11}$	0
C	0	$\frac{2}{11}$	$\frac{7}{11}$	$\frac{2}{9}$
T	$\frac{1}{9}$	0	$\frac{1}{11}$	$\frac{6}{9}$

Example

Kokia yra suma kiekvieno stulpelio? Kodėl?

$S_1 \setminus S_0$	A	G	C	T
A	$\frac{7}{9}$	0	$\frac{1}{11}$	$\frac{1}{9}$
G	$\frac{1}{9}$	$\frac{9}{11}$	$\frac{2}{11}$	0
C	0	$\frac{2}{11}$	$\frac{7}{11}$	$\frac{2}{9}$
T	$\frac{1}{9}$	0	$\frac{1}{11}$	$\frac{6}{9}$

Pavyzdys

40-bazių ilgio protėvio seka S_0 ir palikuonio seka S_1 . Turim sąlyginės tikimybės matricą.

$$S_0 : ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT$$
$$S_1 : ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC$$

$S_1 \setminus S_0$	A	G	C	T
A	7	0	1	1
G	1	9	2	0
C	0	2	7	2
T	1	0	1	6

$S_1 \setminus S_0$	A	G	C	T
A	$\frac{7}{9}$	0	$\frac{1}{11}$	$\frac{1}{9}$
G	$\frac{1}{9}$	$\frac{9}{11}$	$\frac{2}{11}$	0
C	0	$\frac{2}{11}$	$\frac{7}{11}$	$\frac{2}{9}$
T	$\frac{1}{9}$	0	$\frac{1}{11}$	$\frac{6}{9}$

$$P(S_1 = i \mid S_0 = j)$$

Apibrėžimai

$$P_i = P(\text{ i tipo bazė }),\ i \in \Omega = \{A, G, C, T\}$$

$$P_{i|j} = P(S_1 = i \mid S_0 = j),\ i \in \Omega$$

Kiekvienai sekai S_k galimpibrėžti pasiskirstymo vektorių p_k

$$p_k = \begin{bmatrix} P_A & P_G & P_C & P_T \end{bmatrix}^T$$

Pavyzdys

$$p_0 = \begin{bmatrix} P_A & P_G & P_C & P_T \end{bmatrix}^T = \begin{bmatrix} \frac{9}{40} & \frac{11}{40} & \frac{11}{40} & \frac{9}{40} \end{bmatrix}^T$$

$S_0 : ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT$

$S_1 : ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC$

$S_1 \setminus S_0$	A	G	C	T
A	7	0	1	1
G	1	9	2	0
C	0	2	7	2
T	1	0	1	6

$S_1 \setminus S_0$	A	G	C	T
A	$\frac{7}{9}$	0	$\frac{1}{11}$	$\frac{1}{9}$
G	$\frac{1}{9}$	$\frac{9}{11}$	$\frac{2}{11}$	0
C	0	$\frac{2}{11}$	$\frac{7}{11}$	$\frac{2}{9}$
T	$\frac{1}{9}$	0	$\frac{1}{11}$	$\frac{6}{9}$

$$P_{i|j} = P(S_1 = i \mid S_0 = j)$$

Pavyzdys

$$p_1 = \begin{bmatrix} P_A & P_G & P_C & P_T \end{bmatrix}^T = \begin{bmatrix} \frac{9}{40} & \frac{12}{40} & \frac{11}{40} & \frac{8}{40} \end{bmatrix}^T$$

$S_0 : ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT$

$S_1 : ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC$

$S_1 \setminus S_0$	A	G	C	T
A	7	0	1	1
G	1	9	2	0
C	0	2	7	2
T	1	0	1	6

$S_1 \setminus S_0$	A	G	C	T
A	$\frac{7}{9}$	0	$\frac{1}{11}$	$\frac{1}{9}$
G	$\frac{1}{9}$	$\frac{9}{11}$	$\frac{2}{11}$	0
C	0	$\frac{2}{11}$	$\frac{7}{11}$	$\frac{2}{9}$
T	$\frac{1}{9}$	0	$\frac{1}{11}$	$\frac{6}{9}$

$$P_{i|j} = P(S_1 = i \mid S_0 = j)$$

Pavyzdys

Kokia yra sąsaja tarp p_0, p_1 ir $p_{i|j}$?

$$p_0 = \begin{bmatrix} \frac{9}{40} & \frac{11}{40} & \frac{11}{40} & \frac{9}{40} \end{bmatrix}^T$$

$$p_1 = \begin{bmatrix} \frac{9}{40} & \frac{12}{40} & \frac{11}{40} & \frac{8}{40} \end{bmatrix}^T$$

$S_1 \setminus S_0$	A	G	C	T
A	$\frac{7}{9}$	0	$\frac{1}{11}$	$\frac{1}{9}$
G	$\frac{1}{9}$	$\frac{9}{11}$	$\frac{2}{11}$	0
C	0	$\frac{2}{11}$	$\frac{7}{11}$	$\frac{2}{9}$
T	$\frac{1}{9}$	0	$\frac{1}{11}$	$\frac{6}{9}$

$$P_{i|j} = P(S_1 = i \mid S_0 = j)$$

Perėjimo matrica

- Užkoduokime sąlygines tikimybes į matricą.

$$M = \left[P_{i|j} \right] = \begin{bmatrix} P_{A|A} & P_{A|G} & P_{A|C} & P_{A|T} \\ P_{G|A} & P_{G|G} & P_{G|C} & P_{G|T} \\ P_{C|A} & P_{C|G} & P_{C|C} & P_{C|T} \\ P_{T|A} & P_{T|G} & P_{T|C} & P_{T|T} \end{bmatrix}$$

Pavyzdys

$$P_{i|j} = P(S_1 = i \mid S_0 = j)$$

$$M = \begin{bmatrix} P_{A|A} & P_{A|G} & P_{A|C} & P_{A|T} \\ P_{G|A} & P_{G|G} & P_{G|C} & P_{G|T} \\ P_{C|A} & P_{C|G} & P_{C|C} & P_{C|T} \\ P_{T|A} & P_{T|G} & P_{T|C} & P_{T|T} \end{bmatrix} = \begin{bmatrix} \frac{7}{9} & 0 & \frac{1}{11} & \frac{1}{9} \\ \frac{1}{9} & \frac{9}{11} & \frac{2}{11} & 0 \\ 0 & \frac{2}{11} & \frac{7}{11} & \frac{2}{9} \\ \frac{1}{9} & 0 & \frac{1}{11} & \frac{6}{9} \end{bmatrix}$$

$S_1 \setminus S_0$	A	G	C	T
A	$\frac{7}{9}$	0	$\frac{1}{11}$	$\frac{1}{9}$
G	$\frac{1}{9}$	$\frac{9}{11}$	$\frac{2}{11}$	0
C	0	$\frac{2}{11}$	$\frac{7}{11}$	$\frac{2}{9}$
T	$\frac{1}{9}$	0	$\frac{1}{11}$	$\frac{6}{9}$

Pavyzdys

$$p_1 = Mp_0$$

$$\begin{bmatrix} \frac{9}{40} \\ \frac{12}{40} \\ \frac{11}{40} \\ \frac{8}{40} \end{bmatrix} = \begin{bmatrix} \frac{7}{9} & 0 & \frac{1}{11} & \frac{1}{9} \\ \frac{1}{9} & \frac{9}{11} & \frac{2}{11} & 0 \\ 0 & \frac{2}{11} & \frac{7}{11} & \frac{2}{9} \\ \frac{1}{9} & 0 & \frac{1}{11} & \frac{6}{9} \end{bmatrix} \begin{bmatrix} \frac{9}{40} \\ \frac{11}{40} \\ \frac{11}{40} \\ \frac{9}{40} \end{bmatrix}$$

Pavyzdys

$$p_1 = Mp_0$$

$$\begin{bmatrix} \frac{9}{40} \\ \frac{12}{40} \\ \frac{11}{40} \\ \frac{8}{40} \end{bmatrix} = \begin{bmatrix} \frac{7}{9} & 0 & \frac{1}{11} & \frac{1}{9} \\ \frac{1}{9} & \frac{9}{11} & \frac{2}{11} & 0 \\ 0 & \frac{2}{11} & \frac{7}{11} & \frac{2}{9} \\ \frac{1}{9} & 0 & \frac{1}{11} & \frac{6}{9} \end{bmatrix} \begin{bmatrix} \frac{9}{40} \\ \frac{11}{40} \\ \frac{11}{40} \\ \frac{9}{40} \end{bmatrix}$$

Kodėl?

Pavyzdys

$$p_1 = Mp_0$$

$$\begin{bmatrix} \frac{9}{40} \\ \frac{12}{40} \\ \frac{11}{40} \\ \frac{8}{40} \end{bmatrix} = \begin{bmatrix} \frac{7}{9} & 0 & \frac{1}{11} & \frac{1}{9} \\ \frac{1}{9} & \frac{9}{11} & \frac{2}{11} & 0 \\ 0 & \frac{2}{11} & \frac{7}{11} & \frac{2}{9} \\ \frac{1}{9} & 0 & \frac{1}{11} & \frac{6}{9} \end{bmatrix} \begin{bmatrix} \frac{9}{40} \\ \frac{11}{40} \\ \frac{11}{40} \\ \frac{9}{40} \end{bmatrix}$$

Kodėl?

Pavyzdys

$$\begin{bmatrix} ? \\ \\ \\ \\ \end{bmatrix} = \begin{bmatrix} P_{A|A} & P_{A|G} & P_{A|C} & P_{A|T} \\ P_{G|A} & P_{G|G} & P_{G|C} & P_{G|T} \\ P_{C|A} & P_{C|G} & P_{C|C} & P_{C|T} \\ P_{T|A} & P_{T|G} & P_{T|C} & P_{T|T} \end{bmatrix} \begin{bmatrix} P(S_0 = A) \\ P(S_0 = G) \\ P(S_0 = C) \\ P(S_0 = T) \end{bmatrix}$$

Estimation

- Naudojant perėjimo matricą galime įvertinti bazių pasiskirstymo vektorių p_k palikuonių kartų sekose. $S_k, \; k = 1, 2, 3, \ldots$

$$p_k = Mp_{k-1}$$

$$M = \left[P_{i|j} \right] = \left[P(S_k = i \mid S_{k-1} = j) \right]$$

Kas iš to?

- Galime modeliuoti mutacijų dažnius ir jų pokyčius iš kartos į kartą įvertinant ir įvykusias paslėptas mutacijas mutacijas

S0 : ATGTCGCCTGATAATGCC…

S1 : ATGCCGCTTGACAATGCC…

S2 : ATGCCGCGTGATAATGCC…

Faktinis mutacijų skaičius: 5

(dalis yra *paslėptos mutacijos*)

Kam to reikia

- Galime rasti tikrus atstumus tarpsekų

Atstumas tarp diejų sekų

- Filogenezėje reikia apibrėžti atstumą tarp dviejų sekų.
- Jis turi parodyti vidutinį mutacijų dažnį per poziciją, įskaitant nematomas mutacijas.

S0 : ATGTCGCCTGATAATGCC…

⋮

⋮

S : ATGCCGCGTGATAATGCC…

Atstumas tarp dieju sekų

- Matematiniai modeliai
  - Jukes-Cantor
  - Kimura

Jukes-Cantor modelis

Prielaidos

- Bazių pakeitimo (mutacijų) tikimybė išlieka pastovi iš kartos į kartą.
- Mutacijos sekai S_{k-1} virstant seka S_k priklauso tik nuo sekos S_{k-1}.
- Bazė mutuoti į be kurią kitą bazę turi vienodą tikimybę.

Jukes-Cantor modelis

DNR pakaitai

- Ilgą laiką nukleotidas duotoje pozicijoje išlieka tas pats.
- Bet periodiškai tas nukleotidas pasikeičia (visoje populiacijoje)
- Tai vadinama pakaita - kai dominuojantis nukleotidas konkrečioje pozicijoje pakeičiamas kitu.

Jukes-Cantor modelis

Markov'o grandinė pakaitų modeliavimui

- Galimos keturios būklė “a”, “c”, “g”, “t”
- Grandinė veikia bėgant laiko vienetams - būklė gali pasikeisti einant nuo vieno laiko taško prie kito
- Laiko vienetau gali būti pasirenkami bet kaip (tarkim kas 20000 generacijų)

Jukes-Cantor modelis

Prielaidos

- Bazių pakeitimo (mutacijų) tikimybė išlieka pastovi iš kartos į kartą.
- Mutacijos sekai S_{k-1} virstant seka S_k priklauso tik nuo sekos S_{k-1}.
- Bazė mutuoti į be kurią kitą bazę turi vienodą tikimybę.

Jukes-Cantor modelis

- Tarkim kad α yra mažas.
- Mutacijos yra "retos"

Perėjimo matrica **P**

	A	C	G	T
A	$1-\alpha$	$\alpha/3$	$\alpha/3$	$\alpha/3$
C	$\alpha/3$	$1-\alpha$	$\alpha/3$	$\alpha/3$
G	$\alpha/3$	$\alpha/3$	$1-\alpha$	$\alpha/3$
T	$\alpha/3$	$\alpha/3$	$\alpha/3$	$1-\alpha$

Jukes-Cantor modelis

- Tarkim kad α yra mažas.
- Mutacijos yra "retos"

Perėjimo matrica **P**

α yra parametras priklausantis nuo to , ką reiškia laiko vienetas.

Jei laiko vienetas atitinka daug generacijų, α bus didesnis

	A	C	G	**T**
A	1−α	α/3	α/3	α/3
C	α/3	1−α	α/3	α/3
G	α/3	α/3	1−α	α/3
T	α/3	α/3	α/3	1−α

Jukes-Cantor modelis

- α yra parametras priklausantis nuo to , ką reiškia laiko vienetas.
- Jei laiko vienetas atitinka daug generacijų, α bus didesnis
- α - bazės pakeitimo sekos pozicijoje tikimybė per laiko vienetą.
- Pakeitimas reiškia, kad yra 3 galimi pakeitimai (pvz. C → {A,G,T})
- Nesant bazės pakeitimui yra viena galimybė (pvz. C → C)

	A	C	G	T
A	$1-\alpha$	$\alpha/3$	$\alpha/3$	$\alpha/3$
C	$\alpha/3$	$1-\alpha$	$\alpha/3$	$\alpha/3$
G	$\alpha/3$	$\alpha/3$	$1-\alpha$	$\alpha/3$
T	$\alpha/3$	$\alpha/3$	$\alpha/3$	$1-\alpha$

Jukes-Cantor modelis. Supratimui.

- • Tarkim yra perėjimo matrica **P**, ir tikimybių vektorius **v** (eilutė)
- • Ką w = v P atitiktų ?
- • Jei **v** atitinka keturių nukleotidų esamą pasiskirstymą pozicijoje, tai **w** atitiktų nukleotidų pasiskirstymą sekančiu laiko momentu.

Jukes-Cantor modelis. Supratimui.

- Tarkim, kad mes galime rasti tokį vektorių ϕ, kad $\phi P = \phi$
- Jei tikimybių pasiskirstymas yra ϕ, tai jis ir ateityje bus lygus ϕ ateityje.
- Tai vadinama stacionariu Markovo grandinės pasiskirstymu.

Jukes-Cantor modelis. Supratimui.

- Galima patkrinti, kad vektorius $\phi = (0.25, 0.25, 0.25, 0.25)$ tenkina sąlygą $\phi P = \phi$
- Taigi jei pozicija vystysis pakankamai ilgai, bus vienodos galimybės visiems keturiems nukleotidams joje būti.
- Tai labai ilgo laikotarpio spėjimas...bet ar galime išreikšti nukleotidų pasiskirstymą pozicijoje laiko funkcija?

Spektrinė dekompozicija

- Atsiminkit radom ϕ tokį, kad
- $\phi P = \phi$
- Toks vektorius vadinamas **tikriniu vektoriumi,**

  o atitinkanti **tikrinė vertė** yra 1.
- Apskritai jei $u P = \lambda u$ (sakalaras λ), λ yra vadinama tikrine verte ir u yra kairysis tikrinis vektorius P matricai.

Spektrinė dekompozicija

- Panašiai jei P $v^T = \lambda v^T$, tai u yra vadinamas dešniuoju tikriniu vektoriumi.
- Apskritai gali būti daug tikrinių verčių λ_j ir atitinkamų kairiųjų ir dešiniųjų tikrinių vektorių u_j ir v_j.
- Taigi galime parašyti P kaip:

$$P = \sum_j \lambda_j u_j v_j^T$$

	A	C	G	T
A	1-α	α/3	α/3	α/3
C	α/3	1-α	α/3	α/3
G	α/3	α/3	1-α	α/3
T	α/3	α/3	α/3	1-α

- Kadangi musu matrica P yra simetriška, tai kairysis tikrinis vektorius yra lygu dešiniojo tikrinio vektoriaus (tai pačiai tikrinei vertei) transpozicijai. Taigi $u_j=v_j$ ir P galime parašyti kaip:

$$P = \sum_j \lambda_j v_j v_j^T$$

Spektrinė dekompozicija

- Matricos P tikriniai vektoriai v_n ir tikrines vertes λ_n

$\lambda_1 = 1: v_1 = 1/4\ (1\ 1\ 1\ 1)^T$

$\lambda_{2..4} = 1\text{-}4\alpha/3:$

$v_2 = 1/4\ (-1\ -1\ 1\ 1)^T$

$v_3 = 1/4\ (-1\ -1\ -1\ 1)^T$

$v_4 = 1/4\ (1\ -1\ 1\ -1)^T$

Matrica P

	A	C	G	T
A	1-α	α/3	α/3	α/3
C	α/3	1-α	α/3	α/3
G	α/3	α/3	1-α	α/3
T	α/3	α/3	α/3	1-α

$$P = \sum_j \lambda_j v_j v_j^T$$

Spektrinė dekompozicija

- Tuomet, kiekvienam teigiamam sveikam skaičiui t bus teisinga lygybė

$$P^t = \sum_j \lambda^t{}_j v_j v_j{}^T$$

- Kodėl P^t mums turi rūpėti?

•Todėl, kad P^t mums parodo, koks bus pasiskirstymas bazių po t laiko žingsnių.

- Jeigu bazių pasiskirstymas pozicijoje pradžioje buvo lygus vektoriuo v^T, tai tada $P^t v^T$ bus atitinkamas pasiskirstymas po t laiko žingsnių.

Grįžtant prie Jukes-Cantor modelio

• Minėjome, kad ϕ = (.25,.25,.25,.25) yra kairysis P matricos tikrinis vektorius su tikrine verte .

• Kaip jau rodyta, J-C perėjimo matrica P be šios tikrinės vertės turi tikrinę vertę (1-4α/3) ir po matematinių manipuliacijų gautume šios matricos P tokią spektrinę dekompoziciją:

$$P^t = \sum_j \lambda^t{}_j v_j v_j{}^T$$

$$P^t = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix} + (1-\frac{4}{3}\alpha)^t \begin{bmatrix} 3/4 & -1/4 & -1/4 & -1/4 \\ -1/4 & 3/4 & -1/4 & -1/4 \\ -1/4 & -1/4 & 3/4 & -1/4 \\ -1/4 & -1/4 & -1/4 & 3/4 \end{bmatrix}$$

Jukes-Cantor modelis

- Tarkim kad α yra mažas.
- Mutacijos yra "retos"

Perėjimo matrica

$$P(\alpha) = \begin{bmatrix} 1-\alpha & \alpha/3 & \alpha/3 & \alpha/3 \\ \alpha/3 & 1-\alpha & \alpha/3 & \alpha/3 \\ \alpha/3 & \alpha/3 & 1-\alpha & \alpha/3 \\ \alpha/3 & \alpha/3 & \alpha/3 & 1-\alpha \end{bmatrix} \quad p_0 = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}^T$$

Jukes-Cantor modelis

- q(t)=sąlyginė tikimybė, kad bazė esamuoju laiko momentu (t) momentu yra ta pati, kaip ir bazė pradiniu laiko momentu (t=0).

$q(t)$

$$P(\alpha)^t = \begin{bmatrix} \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t \end{bmatrix}$$

Jukes-Cantor modelis

- q(t)=kokia dalis sekos pozicijų neturi stebimų pozicijų.

$q(t)$

$$P(\alpha)^t = \begin{bmatrix} \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t \end{bmatrix}$$

Jukes-Cantor modelis

- p(t)=1-q(t)=dalis sekos pozicijų,kuriose yra stebimos mutacijos.

$q(t)$

$$P(\alpha)^t = \begin{bmatrix} \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t \end{bmatrix}$$

$p(t)=1-q(t)$

Jukes-Cantor modelis

- p(t)=1-q(t)=dalis sekos pozicijų,kuriose yra stebimos mutacijos.

$q(t)$

$$P(\alpha)^t = \begin{bmatrix} \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t \end{bmatrix}$$

$$p(t) = \frac{3}{4} - \frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t$$

$p(t) = 1 - q(t)$

Jukes-Cantor modelis

- p gali būti nustatyta lyginant dvi duotas sekas

$q(t)$

$$P(\alpha)^t = \begin{bmatrix} \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t \\ \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}-\frac{1}{4}\left(1-\frac{4}{3}\alpha\right)^t & \frac{1}{4}+\frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t \end{bmatrix}$$

$$p(t) = \frac{3}{4} - \frac{3}{4}\left(1-\frac{4}{3}\alpha\right)^t$$

$p(t) = 1 - q(t)$

Example from 4.1

Dalis pozicijų, kurose stebimos mutacijos

$$p = \frac{2}{18} = 0.11$$

S0 : ATGTCGCCTGATAATGCC…

S2 : ATGCCGCGTGATAATGCC…

Sekoje stebimos mutacijos: 2

Jukes-Cantor Atstumas

- Žinomam p (ir t), J-C atstumas tarp dviejų sekų S_0 ir S_1 apibrėžiamas taip:

$$d_{JC}(S_0, S_1) = -\frac{3}{4}\ln\left(1 - \frac{4}{3}p\right)$$

$$\begin{array}{ll} S_0 & : \mathrm{ATGTCGCCTGATAATGCC}\ldots \\ & \vdots \\ & \vdots \\ S_1 & : \mathrm{ATGCCGCGTGATAATGCC}\ldots \end{array}$$

Jukes-Cantor Atstumas

- Žinomam p (ir t), J-C atstumas tarp dviejų sekų S_0 ir S_1 apibrėžiamas taip:

$$d_{JC}(S_0, S_1) = -\frac{3}{4}\ln\left(1 - \frac{4}{3}p\right)$$

Kodėl?

$$\begin{array}{ll} S_0 & : \mathrm{ATGTCGCCTGATAATGCC}\ldots \\ & \vdots \\ & \vdots \\ S_1 & : \mathrm{ATGCCGCGTGATAATGCC}\ldots \end{array}$$

Jukes-Cantor Atstumas

α=bazės (simbolio) pakeitimo į kitą dažnis [pakaitų sk. per poziciją per laiko vienetą]

t=laiko vientų skaičius

αt=pakaitų sk. per poziciją įvykęs per *t* laiko vienetų

Jukes-Cantor Atstumas

α=bazės (simbolio) pakeitimo į kitą dažnis [pakaitų sk. per poziciją per laiko vienetą]

t=laiko vientų skaičius

αt=pakaitų sk. per poziciją įvykęs per t laiko vienetų

p(t) - dalis sekos pozicijų,kuriose yra stebimos mutacijos.

$$p = \frac{3}{4} - \frac{3}{4}\left(1 - \frac{4}{3}\alpha\right)^t$$

$$t = \frac{\ln\left(1 - \frac{4}{3}p\right)}{\ln\left(1 - \frac{4}{3}\alpha\right)} \approx \frac{\ln\left(1 - \frac{4}{3}p\right)}{-\frac{4}{3}\alpha}$$

kai α yra mažas

Jukes-Cantor Atstumas

α=bazės (simbolio) pakeitimo į kitą dažnis [pakaitų sk. per poziciją per laiko vienetą]

t=laiko vientų skaičius

αt=pakaitų sk. per poziciją įvykęs per t laiko vienetų

$$p = \frac{3}{4} - \frac{3}{4}\left(1 - \frac{4}{3}\alpha\right)^{t}$$

$$t = \frac{\ln\left(1 - \frac{4}{3}p\right)}{\ln\left(1 - \frac{4}{3}\alpha\right)} \approx \frac{\ln\left(1 - \frac{4}{3}p\right)}{-\frac{4}{3}\alpha}$$

K

$$\boxed{\alpha t \approx -\frac{3}{4}\ln\left(1 - \frac{4}{3}p\right)}$$

Jukes-Cantor Atstumas

Stebimas atstmas tarp sekų d
tai realus atstumas tarp sekų:

$$K \approx -\tfrac{3}{4}\ln\left(1-\tfrac{4}{3}d\right)$$

Tai yra Jukes-Cantor formulė(įsiminkit).

Pavyzdys

Tarkim 40-bazių ilgio pirmykštė ir dabartinė seka yra:

$$S_0 : ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT$$

$$S_1 : ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC$$

$S_1 \setminus S_0$	A	G	C	T
A	7	0	1	1
G	1	9	2	0
C	0	2	7	2
T	1	0	1	6

Pavyzdys

Tarkim 40-bazių ilgio pirmykštė ir dabartinė seka yra:

S_0 : *ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT*

S_1 : *ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC*

$S_1 \setminus S_0$	A	G	C	T
A	7	$\boxed{0}$	$\boxed{1}$	$\boxed{1}$
G	$\boxed{1}$	9	$\boxed{2}$	$\boxed{0}$
C	$\boxed{0}$	$\boxed{2}$	7	$\boxed{2}$
T	$\boxed{1}$	$\boxed{0}$	$\boxed{1}$	6

$$p = \frac{11}{40} = 0.275$$

$$d_{JC} = -\frac{3}{4}\ln\left(1 - \frac{4}{3}\cdot\frac{11}{40}\right) \approx 0.3426$$

Pavyzdys

0.275 stebimų pakaitų per poziciją dažnis.

0.3426 prognozuojamas realiai įvykusių pakaitų per poziciją dažnis

S_0 : $ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT$

S_1 : $ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC$

$S_1 \setminus S_0$	A	G	C	T
A	7	$\boxed{0}$	$\boxed{1}$	$\boxed{1}$
G	$\boxed{1}$	9	$\boxed{2}$	$\boxed{0}$
C	$\boxed{0}$	$\boxed{2}$	7	$\boxed{2}$
T	$\boxed{1}$	$\boxed{0}$	$\boxed{1}$	6

$$p = \frac{11}{40} = 0.275$$

$$d_{JC} = -\frac{3}{4}\ln\left(1 - \frac{4}{3}\cdot\frac{11}{40}\right) \approx 0.3426$$

Pavyzdys

11 stebimų pakaitų

13.7 suprognozuotų pakaitų.

S_0 : $ACTTGTCGGATGATCAGCGGTCCATGCACCTGACAACGGT$

S_1 : $ACATGTTGCTTGACGACAGGTCCATGCGCCTGAGAACGGC$

$S_1 \setminus S_0$	A	G	C	T
A	7	$\boxed{0}$	$\boxed{1}$	$\boxed{1}$
G	$\boxed{1}$	9	$\boxed{2}$	$\boxed{0}$
C	$\boxed{0}$	$\boxed{2}$	7	$\boxed{2}$
T	$\boxed{1}$	$\boxed{0}$	$\boxed{1}$	6

$$p = \frac{11}{40} = 0.275$$

$$d_{JC} = -\frac{3}{4}\ln\left(1 - \frac{4}{3}\cdot\frac{11}{40}\right) \approx 0.3426$$

Jukes-Cantor Atstumas

Jukes-Cantor formulė:

$$\boxed{K \approx -\tfrac{3}{4}\ln\left(1-\tfrac{4}{3}d\right)}$$

Esant mažam d ln(1+x) ≈ x : K ≈ d

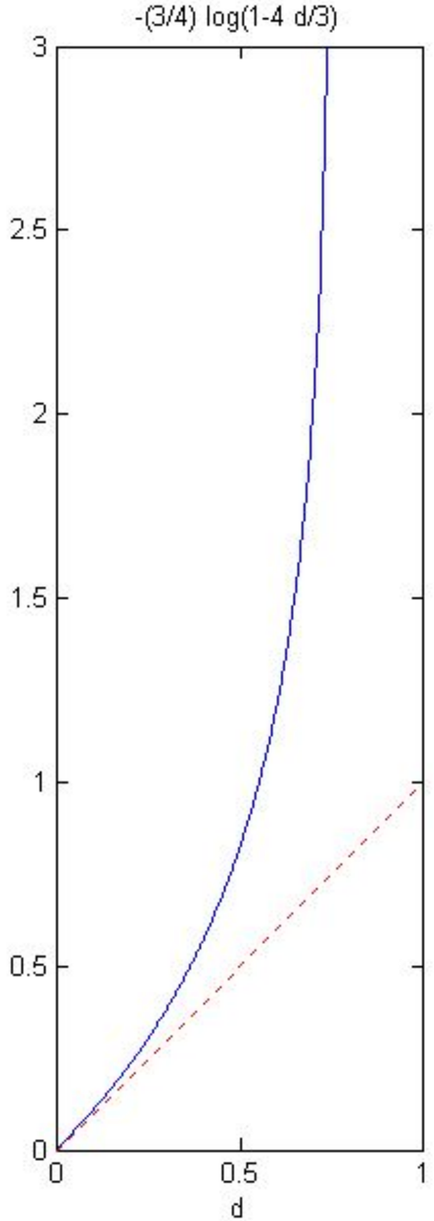
faktinis atstumas ≈ stebimas atstumas

Esant mutacijų įsotinimui: d ↑ ¾ : K →∞

Jei pasikeitė 3/4 lyginamų sekų bazių - atstumas tampa nebenustatomas.

Atsitiktinėm sekoms JC atstumo apskaičiuoti negalima.

$$\boxed{K \approx -\tfrac{3}{4}\ln\left(1-\tfrac{4}{3}d\right)}$$



Jukes-Cantor Atstumas. Dispersija

Jei $K = f(d)$, tuomet $$\delta K = \left(\frac{\partial K}{\partial d}\right)\delta d \Rightarrow \delta K^2 = \left(\frac{\partial K}{\partial d}\right)^2 \delta d^2$$

Taigi $$\mathrm{Var}(K) = \left(\frac{\partial K}{\partial d}\right)^2 \mathrm{Var}(d)$$

Sekų atsiradimas iš pradinės, kurių ilgis n ir pakaitų dažnis *d* yra binminis procesas procesas: $$\mathrm{Prob}(k) = \binom{n}{k} d^k (1-d)^{n-k}$$

ir su tokia dispersija: $\mathrm{Var}(d) = d(1\text{-}d)/n$

O pagal Jukes-Cantor formulę: $$\frac{\partial K}{\partial d} = \frac{1}{1 - \frac{4}{3}d}$$

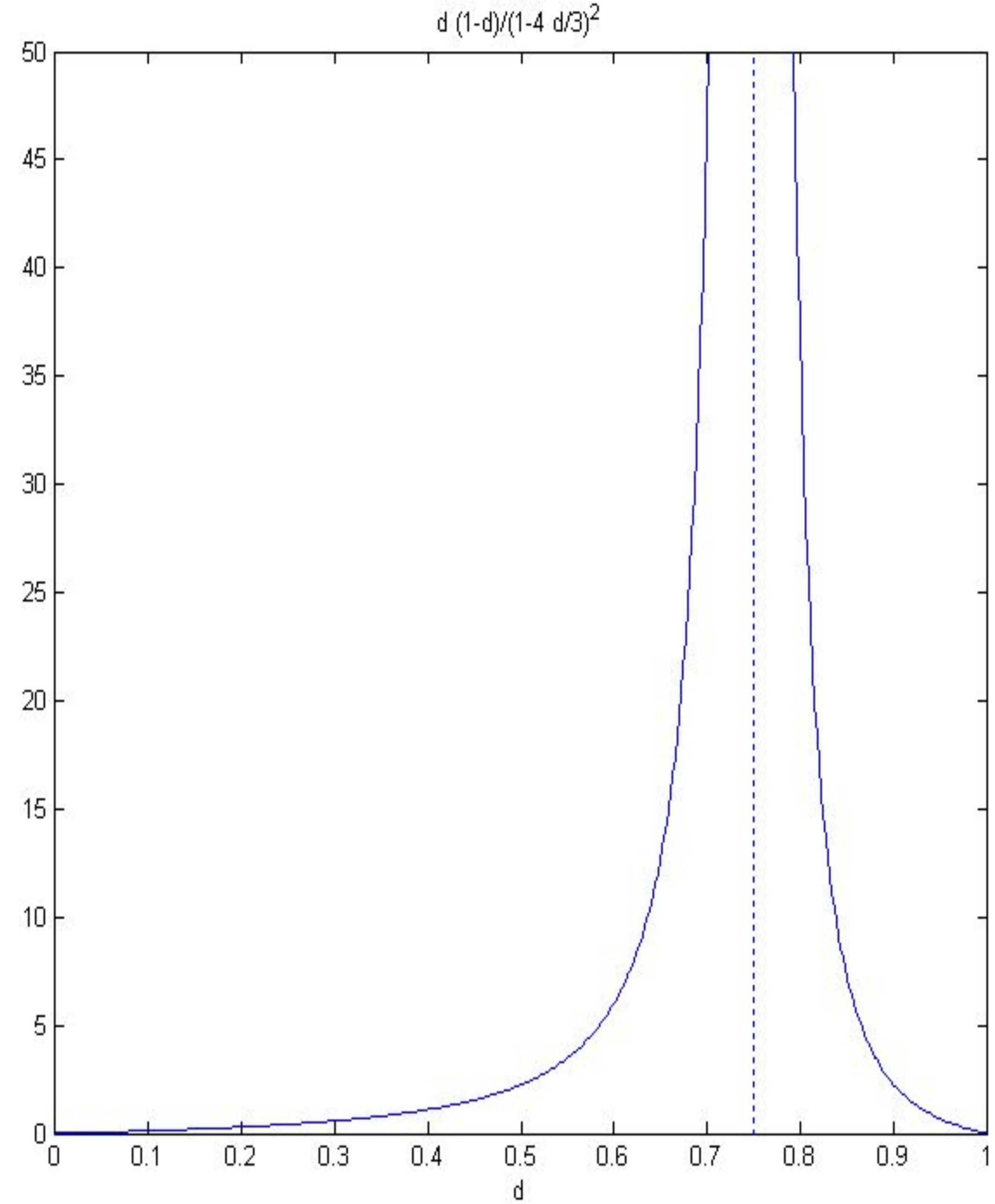
Jukes-Cantor Atstumas. Dispersija

dispersija: $\mathrm{Var}(d) = d(1\text{-}d)/n$

O pagal Jukes-Cantor formulę: $$\frac{\partial K}{\partial d} = \frac{1}{1 - \frac{4}{3}d}$$

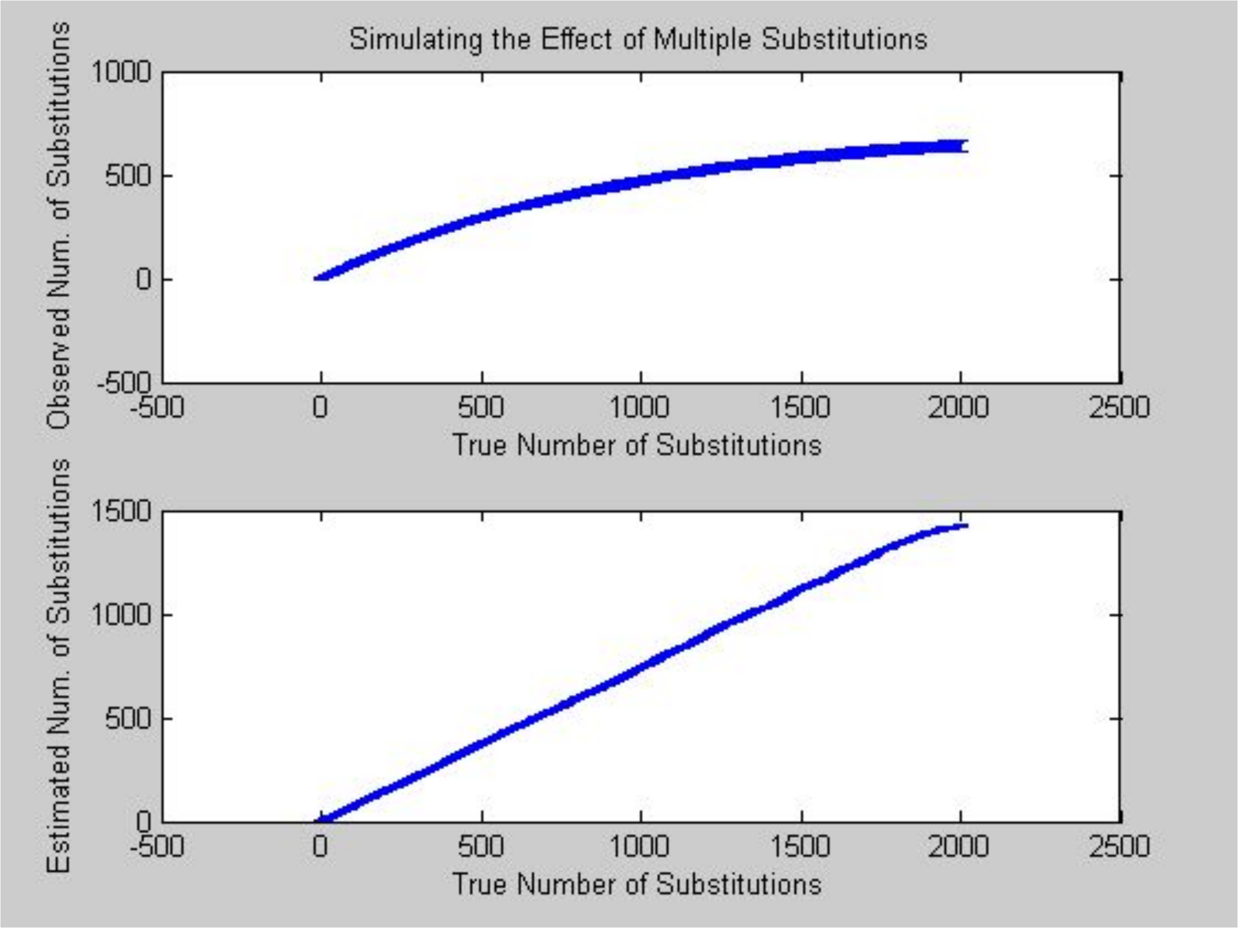
Taigi: $$\boxed{\mathrm{Var}(K) \approx \frac{d(1-d)}{n(1-\frac{4}{3}d)^2}}$$

Var(*K*)



Jukes-Cantor Atstumas.

1000 bp ilgio sekos evoliucijos modeliavimas. Pakaitų dažniai pagal JC prielaidas.



Kimura 2-parametrų atstumas

- Ne visi pakaitai vienodai tikėtini.
- Tranzicija jos tikimybė sekos pozicijoje (G↔A ir T↔C) yra α
- Transversijos tikimybė (G↔T, G↔C, A↔T, and A↔C) sekos pozicijoje yra β

Kimura 2-parametrų atstumas

Atitinkamo Markovo proceso pakaitų matrica:

$$P(\alpha,\beta)=\begin{bmatrix} -\alpha-2\beta & \alpha & \beta & \beta \\ \alpha & -\alpha-2\beta & \beta & \beta \\ \beta & \beta & -\alpha-2\beta & \alpha \\ \beta & \beta & \alpha & -\alpha-2\beta \end{bmatrix}$$

Kimura 2-parametrų atstumas

Pot generacijų pakaitų tikimybė bus:

$$P(t) = P_{K2P}{}^{t}$$

Reikia nustatyti P(t):

tikrines vertes $\{\lambda_i\}$

ir **tikrinius vektorius** $\{\boldsymbol{v}_i\}$

Kimura 2-parametrų atstumas

Spektrinė dekompozicija $M(t)$:

$$P_{K2P}{}^{t} = \sum_i \lambda_i^t \mathbf{v}_i \mathbf{v}_i^{\mathrm{T}}$$

$$P^n = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix} + (1-4\beta)^n \begin{bmatrix} 0.25 & 0.25 & -0.25 & -0.25 \\ 0.25 & 0.25 & -0.25 & -0.25 \\ -0.25 & -0.25 & 0.25 & 0.25 \\ -0.25 & -0.25 & 0.25 & 0.25 \end{bmatrix}$$

$$+ (1-2(\alpha+\beta))^n \begin{bmatrix} 0.5 & -0.5 & 0 & 0 \\ -0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 \\ 0 & 0 & -0.5 & 0.5 \end{bmatrix}$$

Kimura 2-parametrų atstumas

Nustatykite kokia dalis bus tranzicijų per poziciją per t generacijų : $\boldsymbol{P(t)}$

Nustatykite kokia dalis bus transversijų per poziciją per t generacijų : $\boldsymbol{Q(t)}$

Atstumas tarp sekų: $\boldsymbol{K \approx -\tfrac{1}{2} \ln(1-2P-Q) - \tfrac{1}{4} \ln(1-2Q)}$

Pakaitų dalis $\boldsymbol{d = P + Q \rightarrow}$ ***Jukes-Cantor***

Kimura 2-parametrų atstumas

$$P^n = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix} + (1-4\beta)^n \begin{bmatrix} 0.25 & 0.25 & -0.25 & -0.25 \\ 0.25 & 0.25 & -0.25 & -0.25 \\ -0.25 & -0.25 & 0.25 & 0.25 \\ -0.25 & -0.25 & 0.25 & 0.25 \end{bmatrix}$$

$$+ (1-2(\alpha+\beta))^n \begin{bmatrix} 0.5 & -0.5 & 0 & 0 \\ -0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 \\ 0 & 0 & -0.5 & 0.5 \end{bmatrix}$$

Kiti evoliucijos modeliai

- Atskiria kitus tranzicijos transvercijos tipus - naudoja daugiau parametrų
- ...daugybė kitų metodų. Naudojama net 15-os parametrų
- Esmini trūkumas egzistuojančių metodų yra tai, kad remiamasi prielaida, kad simetrinių mutacijų tikimybės yra vienodos

$$\mathrm{prob}(A \to T) = \mathrm{prob}(T \to A)$$

- Gana daug stiprių įrodymų, kad ši prielaida yra klaidinga....
- Ateities iššūkiai...

Pavyzdys. Ar neandartaliečiai yra tarp mūsų.

- 206 mtDNA molekulės iš H. sapiens.
- 2 mtDNA molekulės iš H. neanderthaliensis bei 1856 paskiri fragmentai.
- Visi 208 mėginiai iš GenBanko.

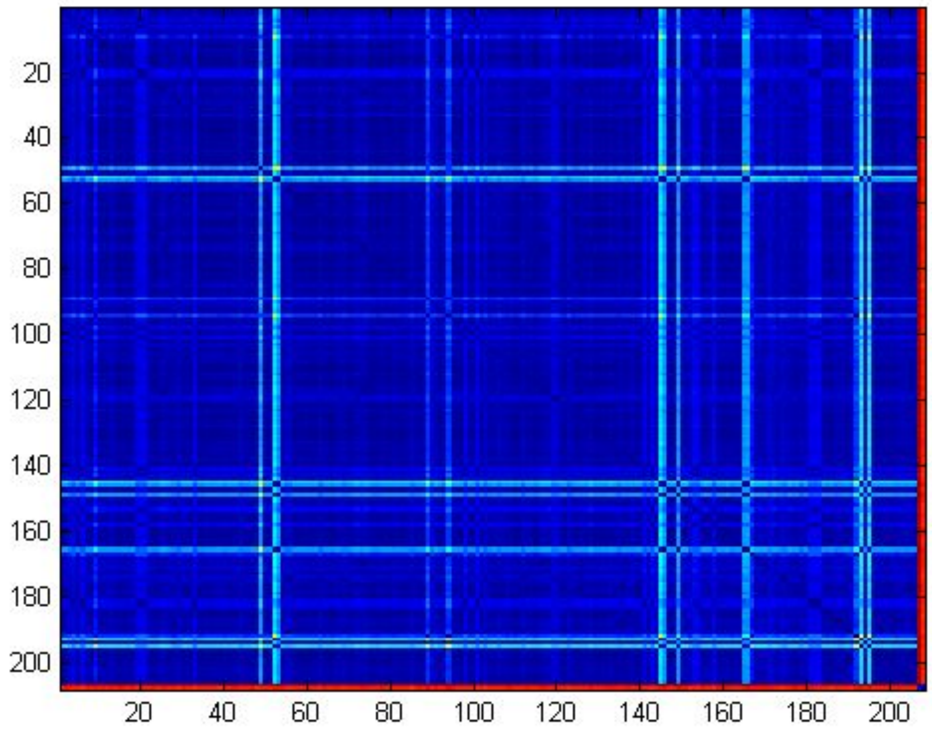
- Homologinė seka 800 bp rasta visuose mėginiuose.

Pavyzdys. Ar neandartaliečiai yra tarp mūsų?

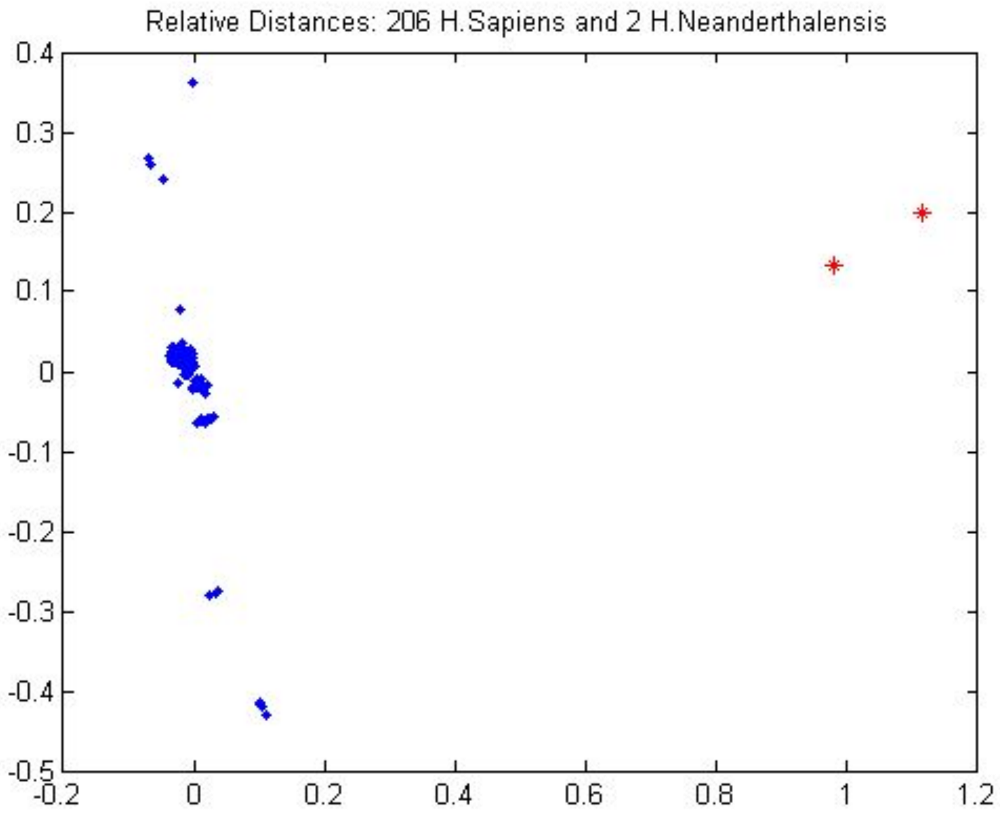
- 206 mtDNA molekulės iš H. sapiens.
- 2 mtDNA molekulės iš H. neanderthaliensis bei 1856 paskiri fragmentai.
- Visi 208 mėginiai iš GenBanko.

- Homologinė seka 800 bp rasta visuose mėginiuose.

Pavyzdys. Ar neandartaliečiai yra tarp mūsų?



Pavyzdys. Ar neandartaliečiai yra tarp mūsų?



Pavyzdys. Ar neandartaliečiai yra tarp mūsų?