

Algoritminen bioinformatiikka

Veli Mäkinen
Tietojenkäsittelytieteen laitos

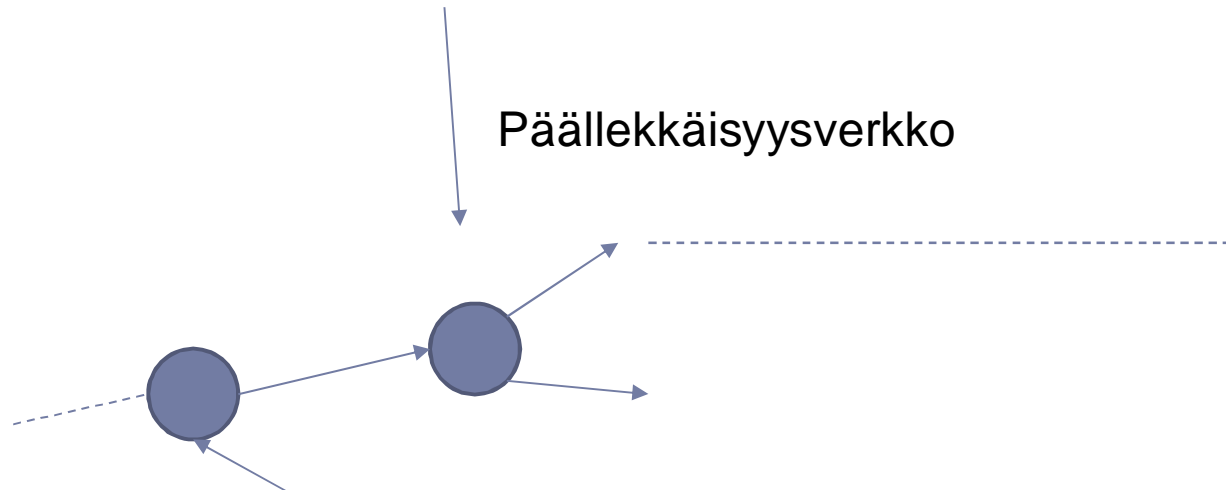
Sisältö

- ▶ **Esimerkkiongelmia**
 - ▶ Data-analyysiä
 - ▶ I) Genomin koostaminen lyhyistä sekvenssipalasista
 - ▶ II) Vaihtoehtoinen silmukointi, isoformien ilmentymistasot
 - ▶ Laskennallista molekyylibiologiaa
 - ▶ III) Motiivit
 - ▶ IV) Genomin uudelleenjärjestymiset, geeniklusterit
 - ▶ V) Evoluutiopuut
 - ▶ Jotain siltä väliltä:
 - ▶ VI) Merkkijonoydinfunktiot
- ▶ **Mitä on ei-algoritminen bioinformatiikka?**



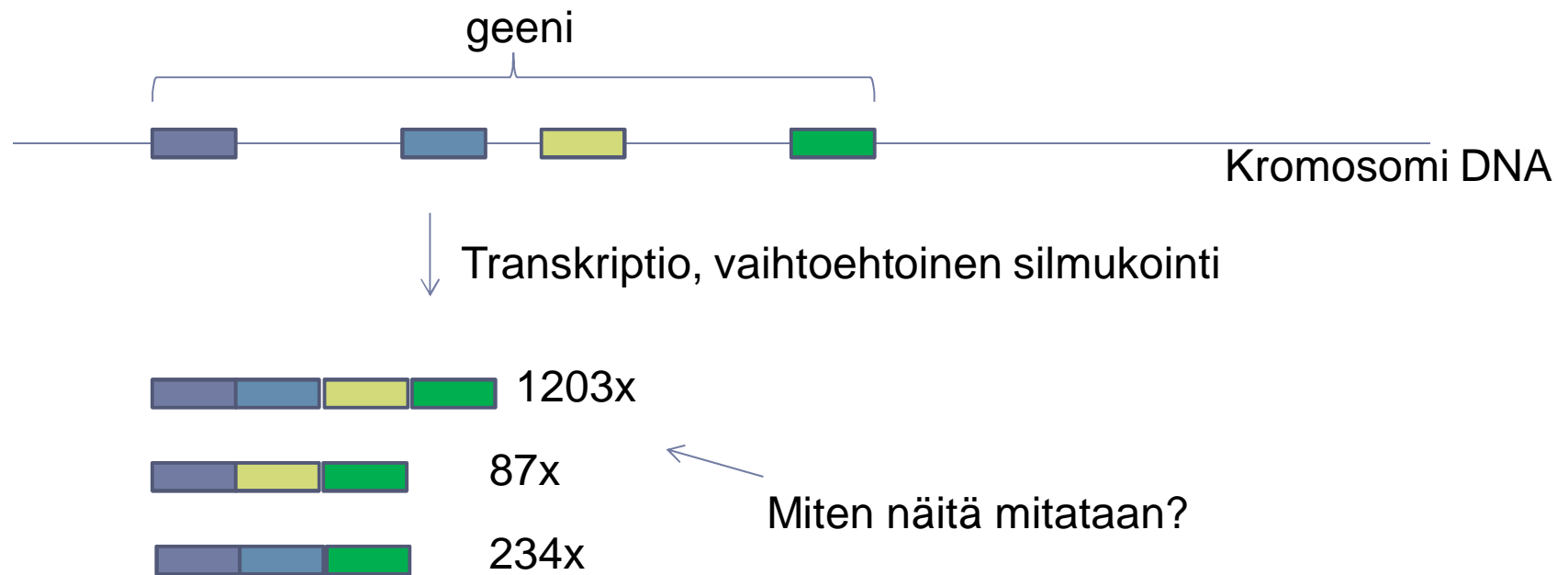
I) Genomin koostaminen

← lineaarisen kromosomin DNA
AGCATCATCGACAGTCGATCATCGAGCGCGCGCGCGTATA
ACAGTCGATCAT
GCATGATCGAC

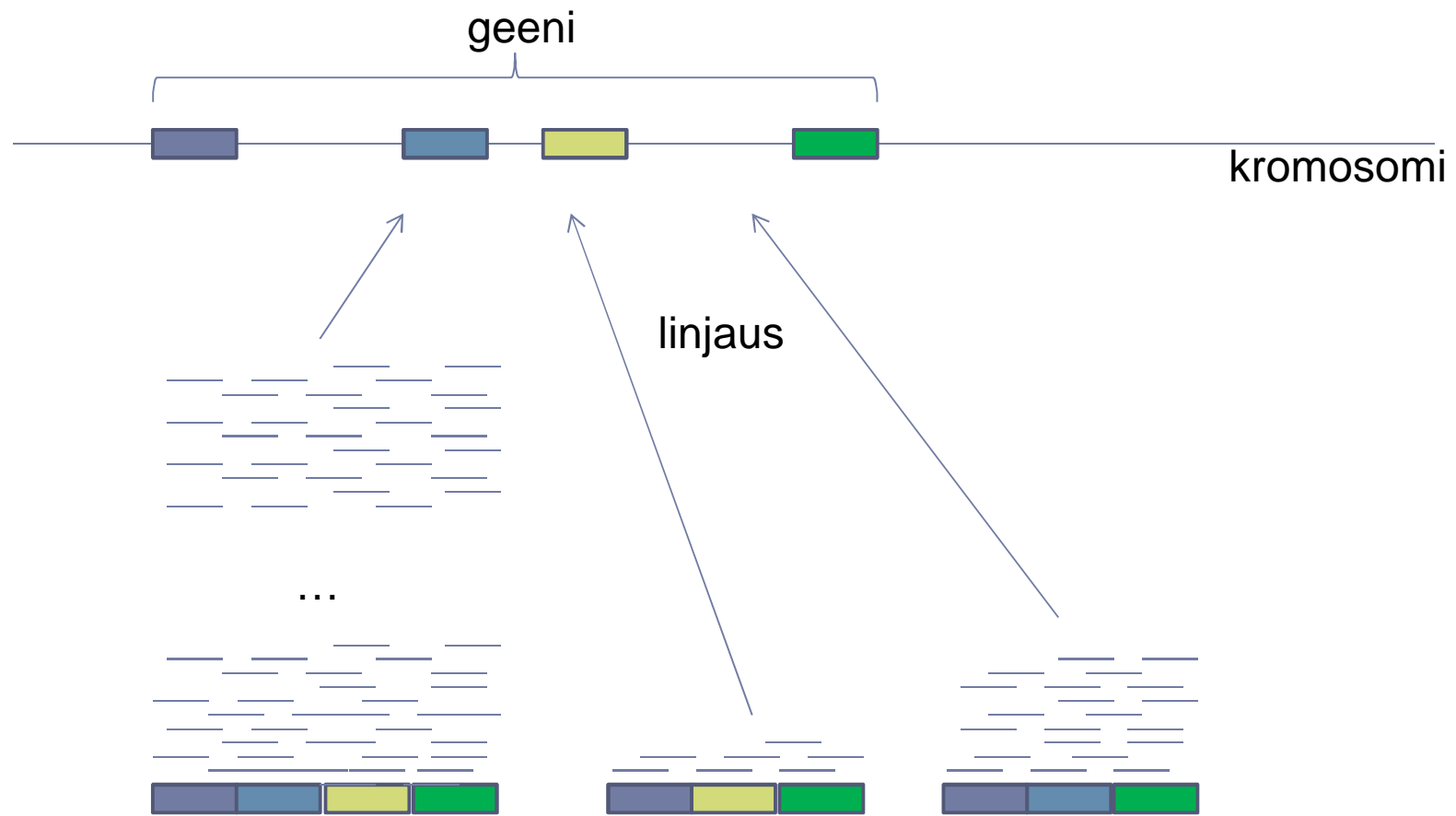


- Ilman virheitä, täydellä peitolla, on olemassa polku joka vastaa genomia (kromosomi:
... Mutta on paljon muitakin polkuja
... Kaksijuosteisuus, diploidisuus, sekvensointivirheet, useat kromosomit

II) Silmukointiverkot, transkriptomiikka

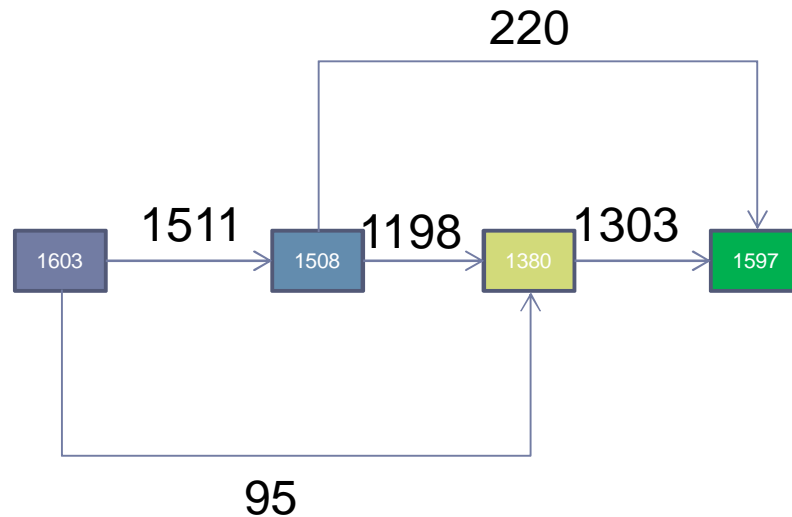


II) RNA-sekvensointi, hyppivät linjaukset

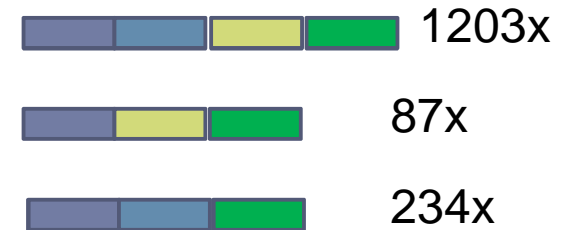


→ Biological Sequence Analysis, 5 op

II) Silmukointiverkko

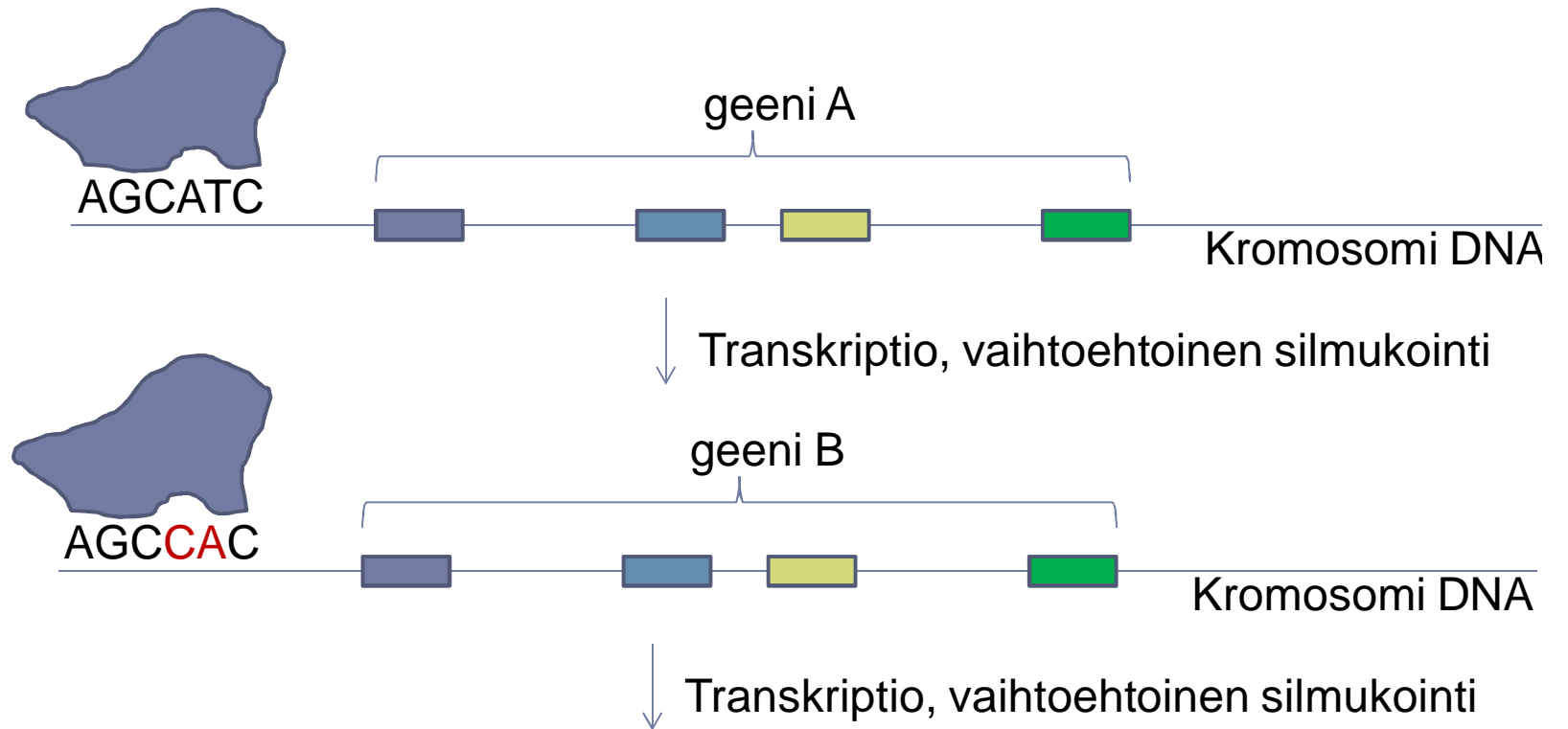


Etsi polut jotka parhaiten selittävät verkon, kuten:



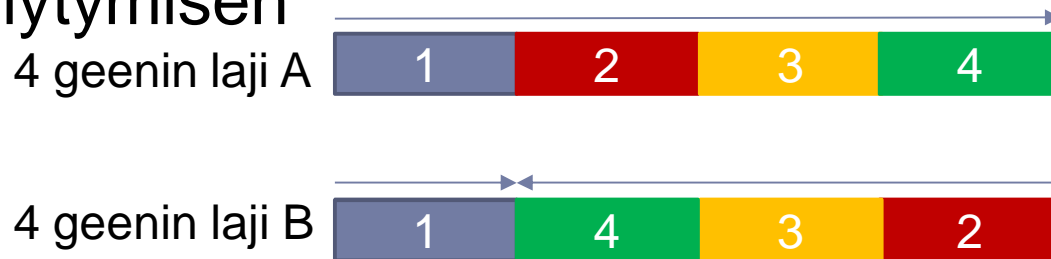
- ▶ Ongelma voidaan palauttaa minimipainoisen verkkovuon laskemiseen (minimum cost flow). Tämä on klassinen kombinatorisen optimoinnin ongelma (Ford & Fulkerson, 56, ...)

III) Motiivit



IV) Genomin uudelleenjärjestymiset

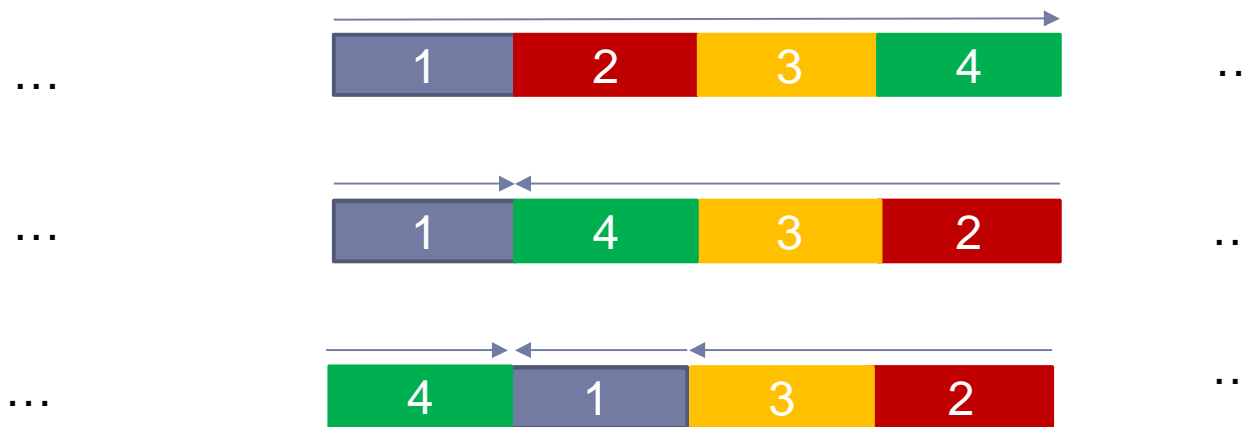
- ▶ Evoluutiossa tapahtuu lajin sisällä pieniä muutoksia, mutta isommat muutokset aiheuttavat lajien eriytymisen



- ▶ Lajien välinen etäisyys: montako suurta muutosta on tapahtunut?
- ▶ Paljon syvällistä algoritmiikkaa on kehittynyt tämän aiheen ympärille!

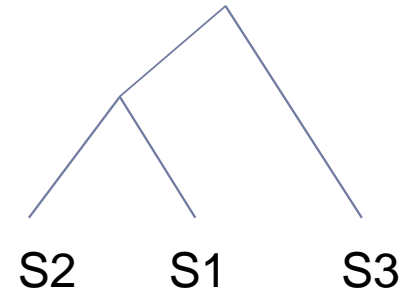
IV) Geeniklusterit

- ▶ Löytyykö usealle lajille yhteisiä alueita, joissa samat geenit mutta eri järjestyksessä?
- ▶ Ehkä näillä geeneillä on jokin yhteinen tehtävä



V) Evoluutiopuut

	S1	S2	S3
S1	0	2	4
S2	2	0	4
S3	4	2	0



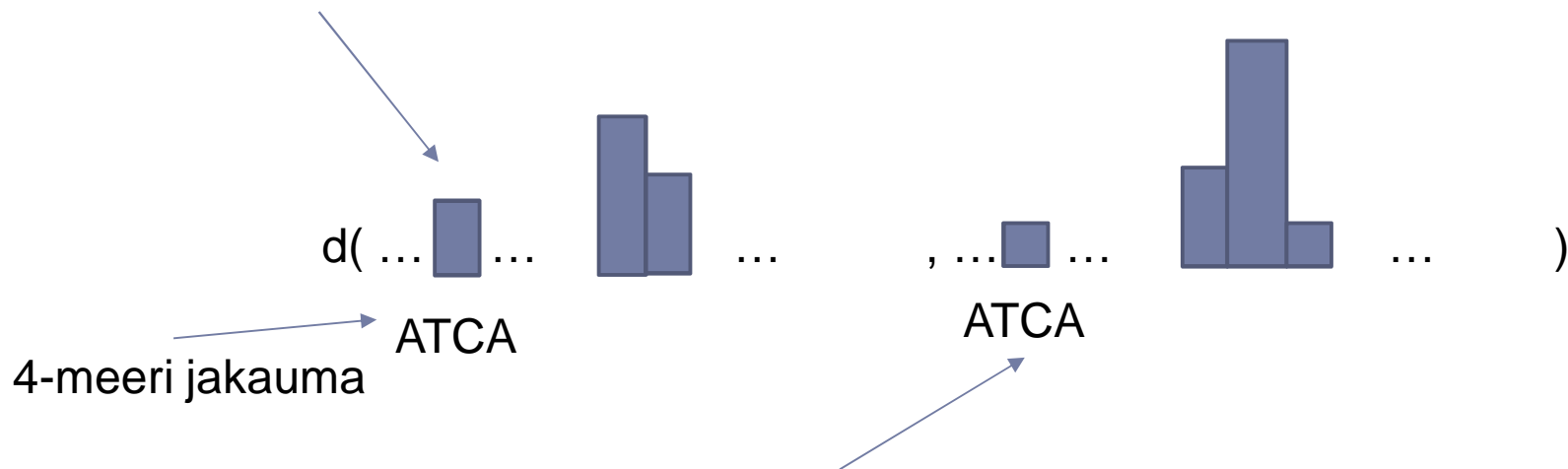
- ▶ Helppoa jos etäisyydet ovat tarkkoja... mutta haastavaa käytännössä
- ▶ Paljon kirjallisuutta sekä algoritmiikassa että tilastotieteessä



VI) Merkkijonoydinfunktiot

AGCATCATCGACAGTCGATCATCGAGCGCGCGCGCGTATA

GCATCATCGAC



TATATATGATCAGCTACTAATACGCACGAAAACGTTATTATAT

TATATGATCAGC

Bioinformatiikkaa

Molekyylibiologia	Tilastotiede	Algoritmit	Ohjelmistot / hajautus
Yleiskuva koko kentästä	Todennäköisyysmallinnus ilmiöistä	Kombinatorinen mallinnus	Dataformaatit, tietokannat
"best practices"	Iteratiiviset optimointimenetelmät	Räätälöidyt algoritmit kuhunkin ongelmaan	Kokonaisanalyysin skaalaus laskentaklusterilla
Perustason tietämys menetelmistä	Ymmärrys eri tilastollisten menetelmien soveltuvuudesta eri ongelmiin	Ymmärrys räätälöidä algoritmeja eri ongelmiin	Ohjelmointi hajautetussa ympäristössä

Koneoppiminen

Laskennallinen biologia, teorettinen biologia, matemaattinen biologia,...



Suo siellä, vetelä täällä

