Postoje li osetljivi delovi u ljudskom genomu?

Kombinatorni algoritmi

Bioinformatics Algorithms:

an Active Learning Approach

~Poglavlje 6~

Pregled

- **Transformacija čoveka u miša**
- Sortiranje po promenama
- Teorema o prekidnoj tački
- Preuređivanje u multihromozomalnim genomima
- Problem rastojanja 2-prekida
- Grafovi prekidnih tačaka
- Teorema o rastojanju 2-prekida

Preuređenja genoma

- Tokom evolucije koja je trajala milionima godina nastajale su različite biljne i životinjske vrste
- U dalekoj prošlosti, neke današnje vrste su imale zajedničkog pretka
- S obzirom da svaka vrsta ima jedinstveni DNK, razvoj vrsta je podrazumevao promenu DNK zajedničkog pretka
- Od promena DNK do sada smo pominjali različite mutacije: supstitucije (zamena jednog nukleotida drugim), insercije i delecije
- Pored takvih, postoje i druge promene u genomu odnosno u DNK sekvenci koje se mogu dogoditi, preneti na potomke i tako omogućiti stvaranje nove vrste
- Ove promene zovemo *preuređenjima genoma* (eng. *genome rearrangements*)

Preuređenja genoma

- Ilustrujmo prethodno razmatranje na primeru čoveka i miša
- Obe ove vrste pripadaju sisarima i pre oko 75 miliona godina su počele da se odvajaju od zajedničkog pretka
- Iako su danas ove vrste sasvim različite, ostala je sličnost u njihovim genomima
- Naime, ako bismo isekli 23 ljudska hromozoma na 280 delova i dobijene fragmente DNK složili po drugačijem redosledu, dobili bismo mišji genom

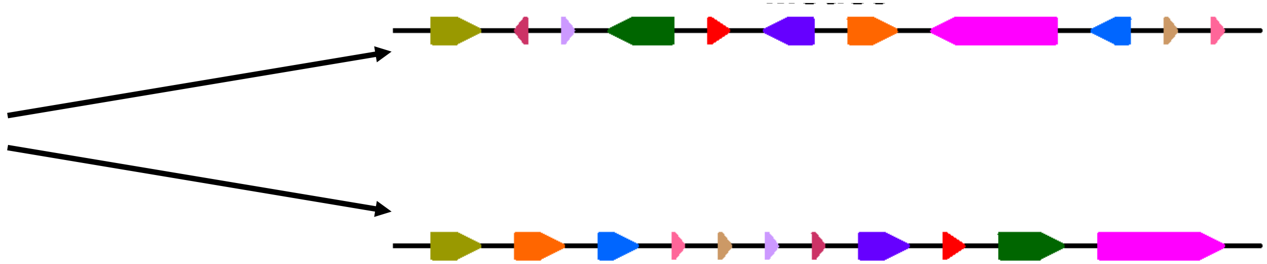
Transformacija čovek-miš

- Kako bismo pojednostavili poređenje dva genoma, ograničićemo se na poređenje samo jednog hromozoma miša i čoveka, i to hromozoma X
- Ovaj hromozom je jedan od dva hromozoma koji određuju pol i tokom evolucije je zadržao skoro sve gene
- Slični geni, često i po nekoliko stotina njih, grupišu se često jedan pored drugog na hromozomu i zadržavaju ovaj redosled u genomima različitih vrsta
- Ovakvu grupu gena nazivamo *blokom sintenije*

Preuređenje genoma

Nepoznati zajednički
predak čoveka i miša
~ pre 75 miliona godina

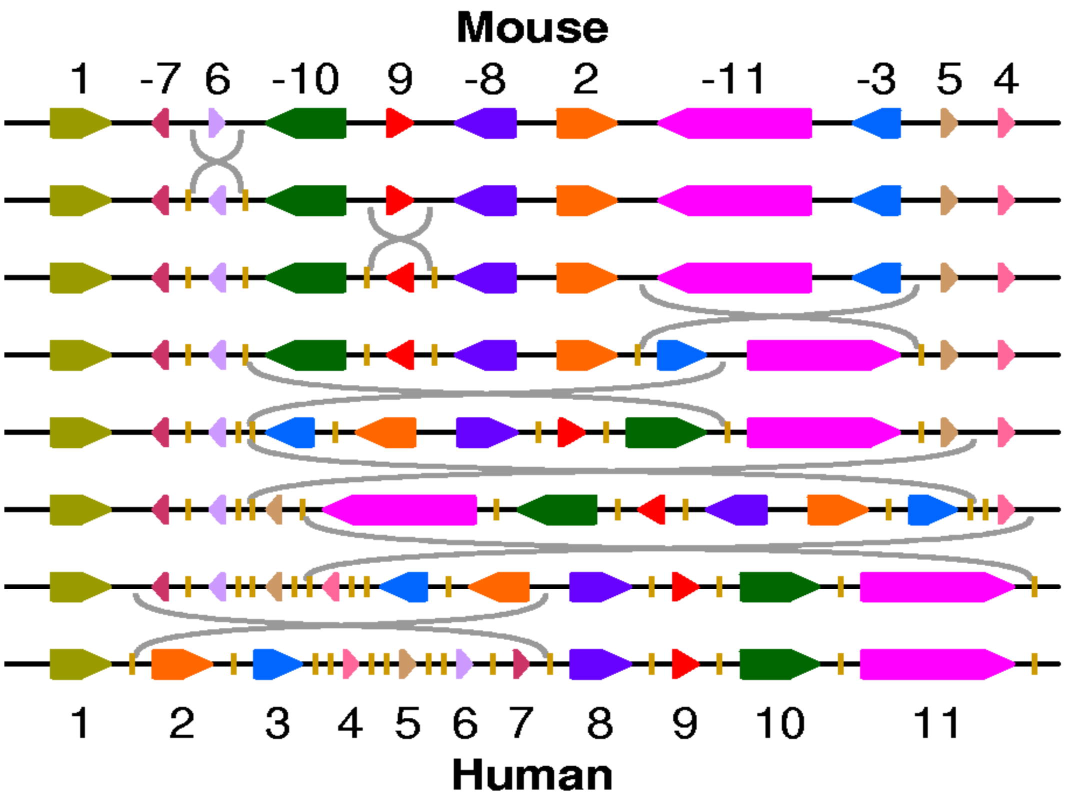
miš (X hromozom)



čovek (X hromozom)

- Koji blokovi genoma su slični i kako da ih nađemo?
- Kakav bi bio evolucioni scenario za transfomisanje jednog genoma u drugi?

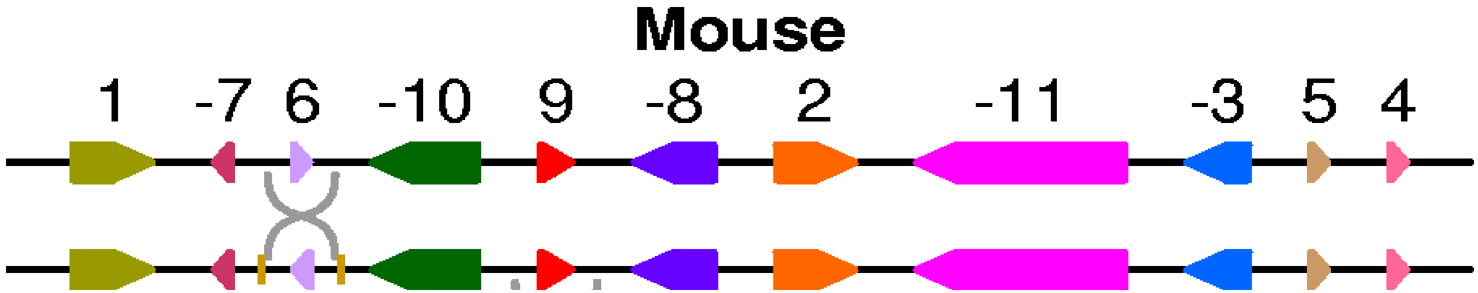
Transformacija miša u čoveka



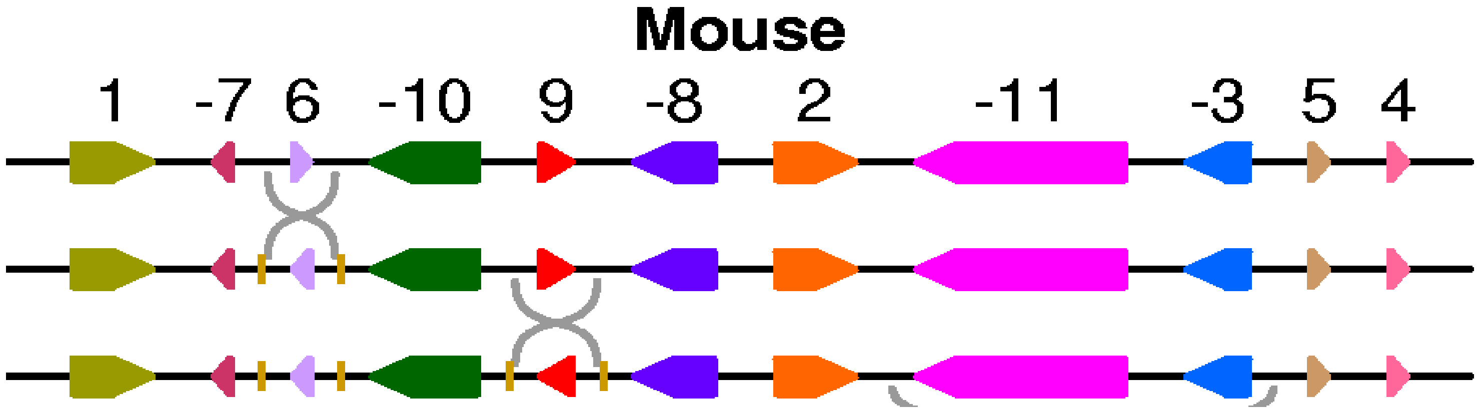
- Svaki blok sintenije je označen celim brojem a promena znaka ukazuje na to da je taj segment DNK tokom preuređenja obrnut (na primer, AAGTAG nakon obrtanja postaje GATGAA)

- Promena između dva susedna koraka podrazumeva da se nekoliko susednih blokova sintenije obrne.

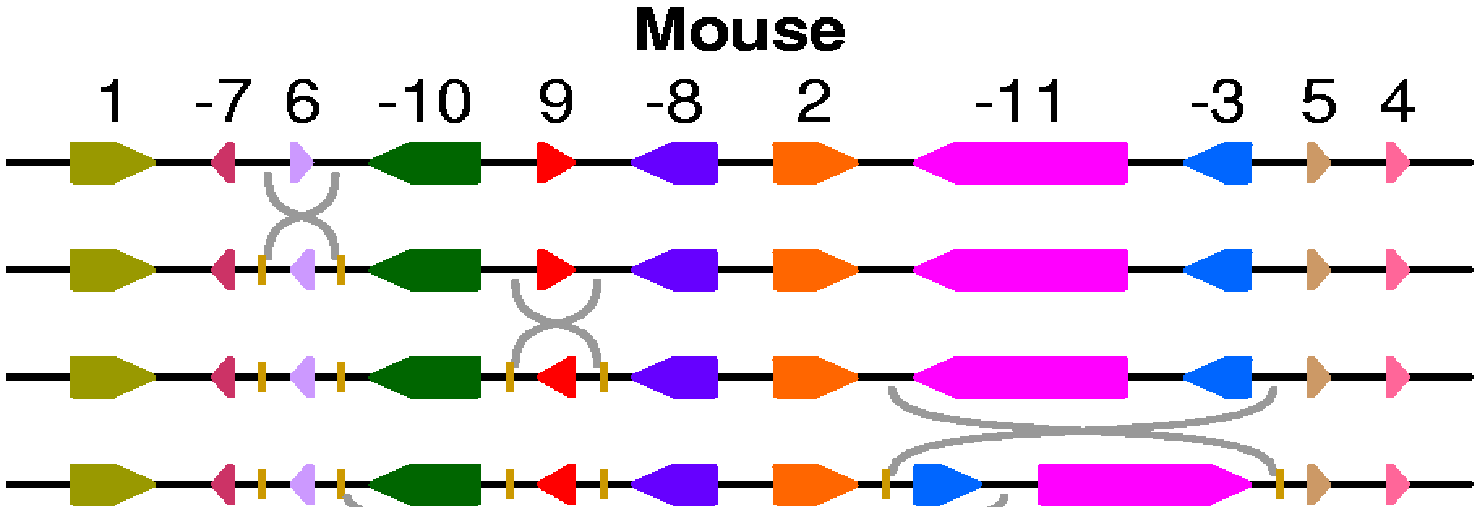
Niz promena



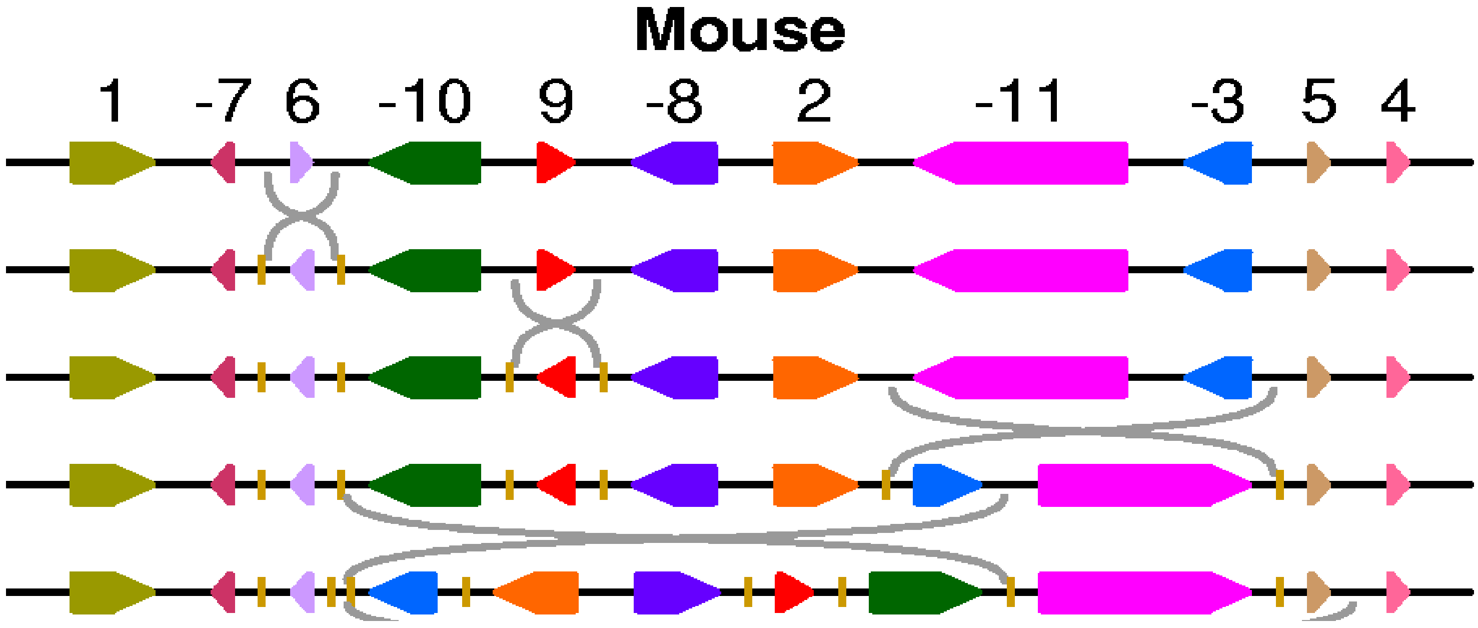
Niz promena



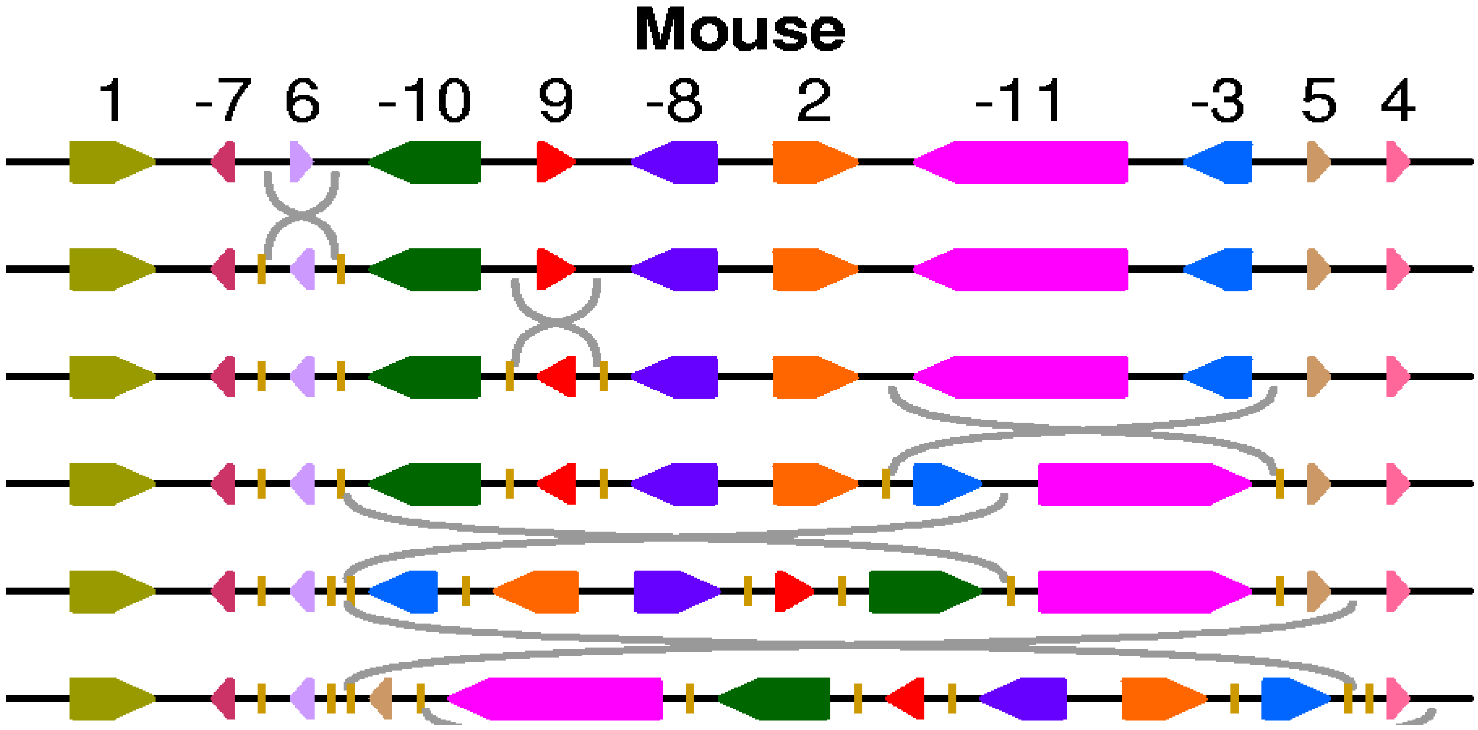
Niz promena



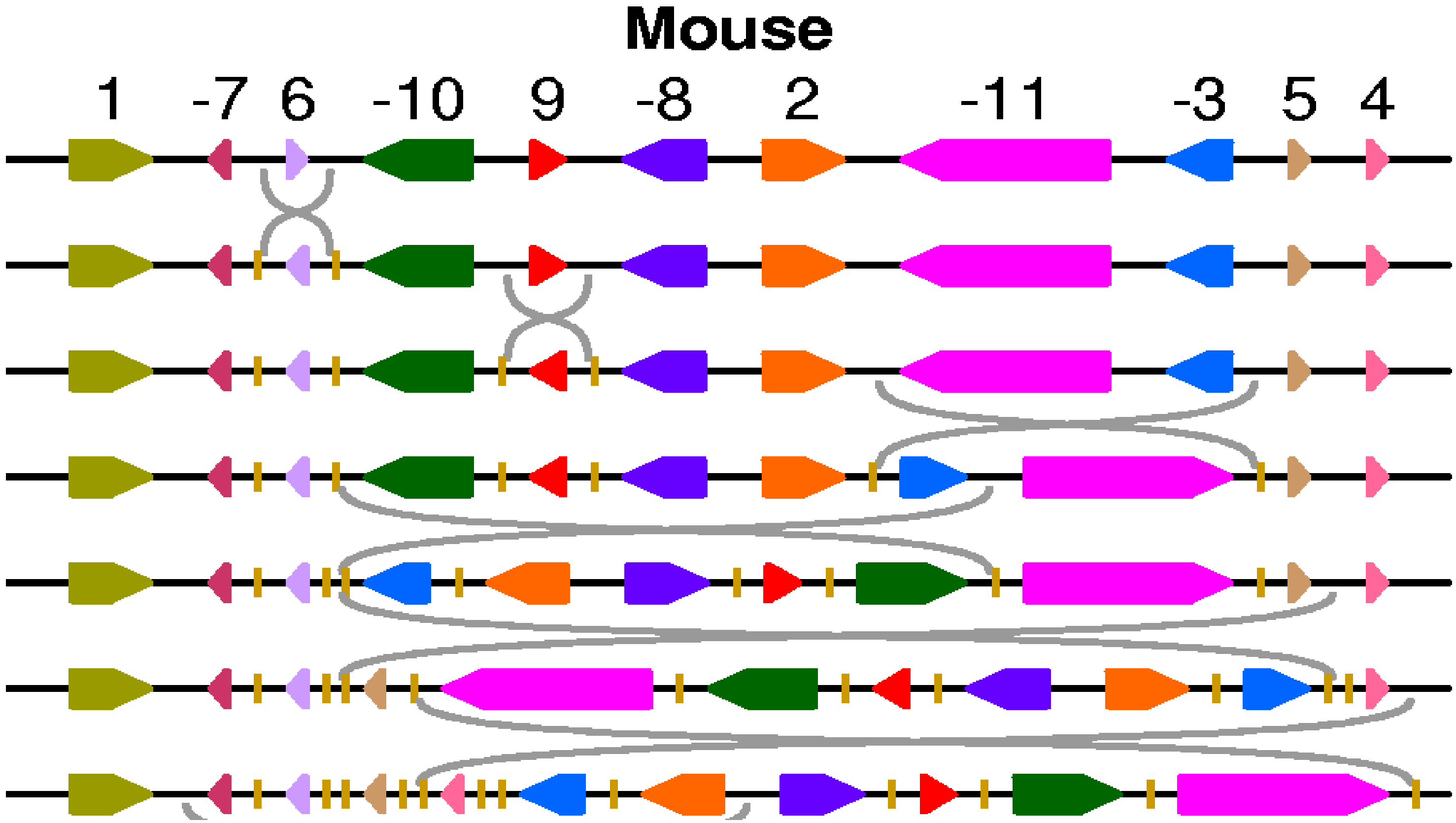
Niz promena



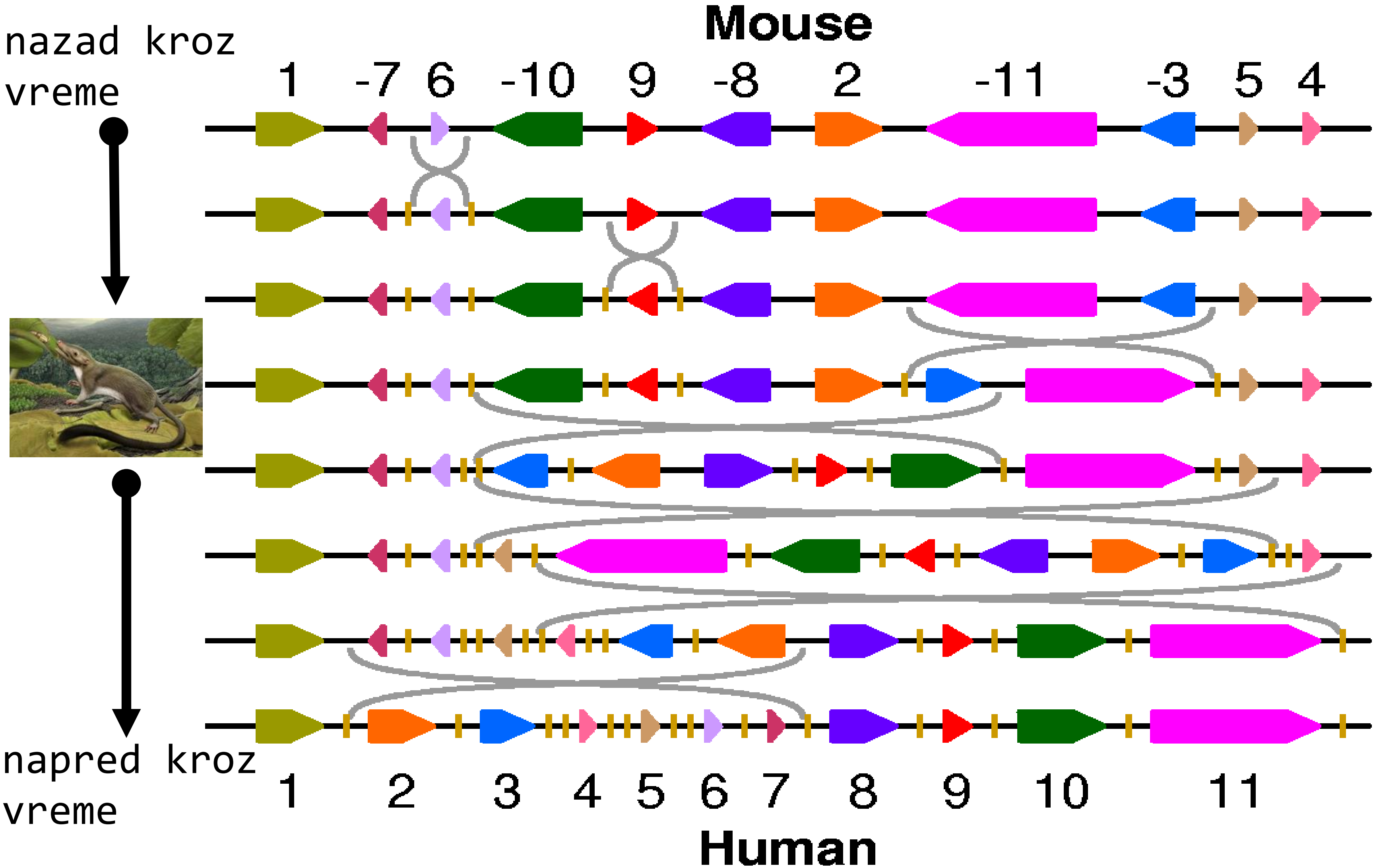
Niz promena



Niz promena



Završena transformacija čovek-miš!



Transformacija čovek-miš

- Ovo je samo jedan od velikog broja mogućih scenarija preuređenja genoma
- Nije poznato da li je bilo ukupno 7 promena, a čak i da ih je bilo tačno toliko, da su išle baš ovim redom
- Kako bismo izgradili matematički model za analizu preuređenja genoma, pretpostavićemo da je tokom evolucije bilo *najmanje moguće promena* između genoma miša i čoveka, odnosno između zajedničkog pretka i svake od ove dve vrste.
- Razmotrimo pitanje *koji je minimalan broj koraka neophodan za transformaciju jednog genoma u drugi*.

Transformacija čovek-miš

- Formalno, svaki genom možemo prestaviti kao jednu permutaciju svojih blokova sintenije uz dodavanje znaka za njihovo usmerenje

- Tako se prikazani genomi mogu predstaviti na sledeći način:

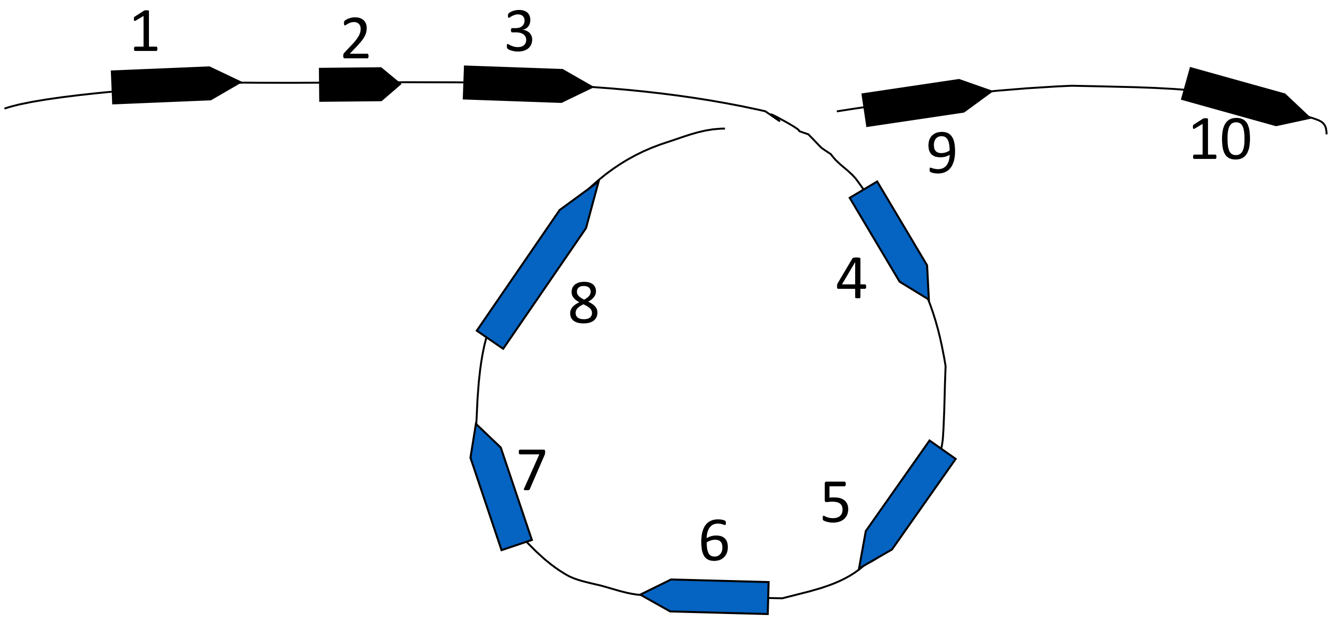
Mouse: (+1 -7 +6 -10 +9 -8 +2 -11 -3 +5 +4)

Human: (+1 +2 +3 +4 +5 +6 +7 +8 +9 +10 +11)

Pregled

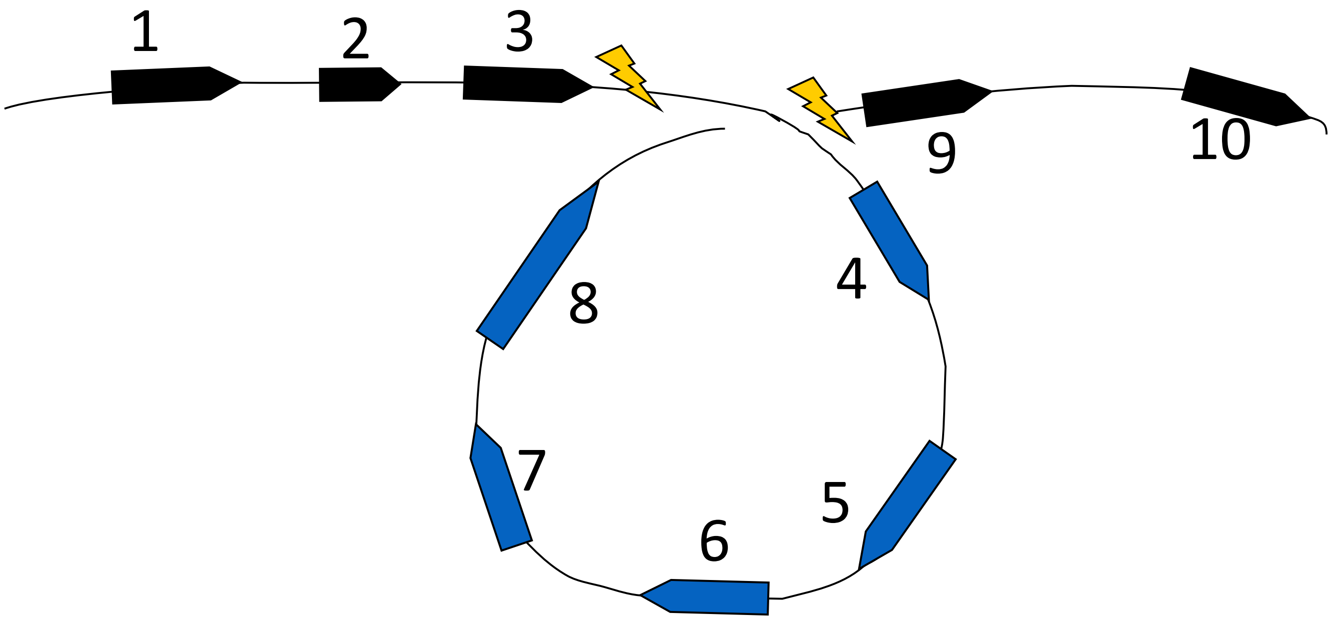
- Transformacija čoveka u miša
- **Sortiranje po promenama**
- Teorema o prekidnoj tački
- Preuređivanje u multihromozomalnim genomima
- Problem rastojanja 2-prekida
- Grafovi prekidnih tačaka
- Teorema o rastojanju 2-prekida

Promene



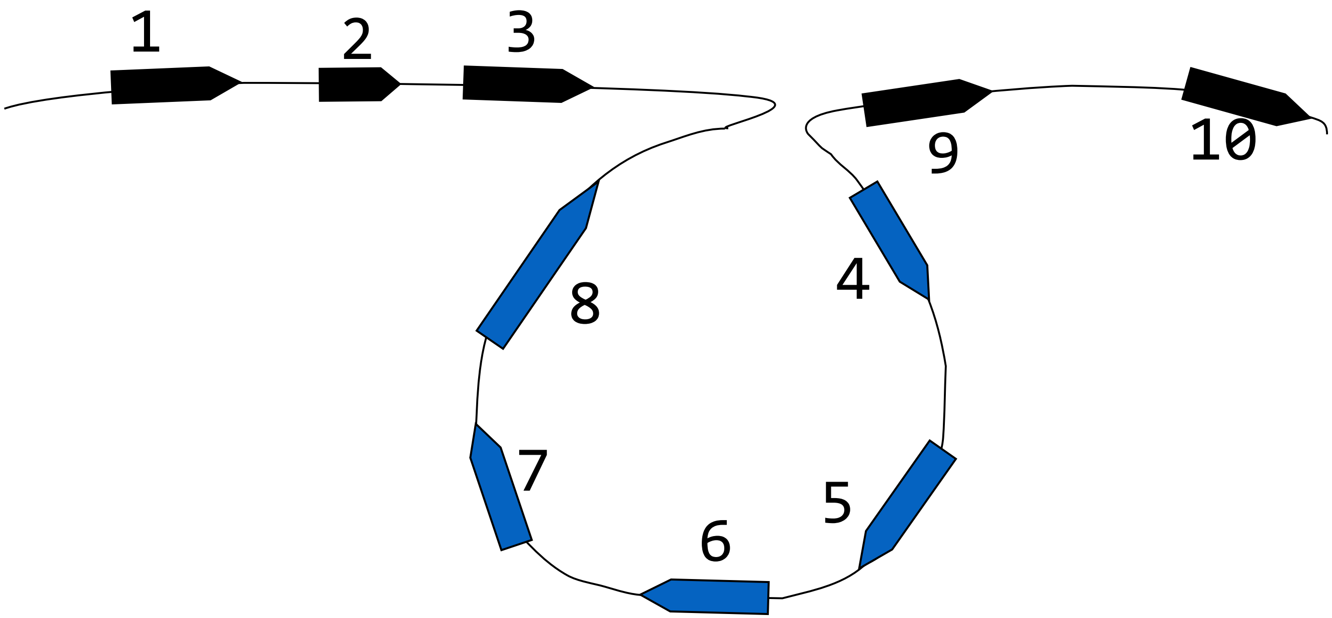
+1 +2 +3 +4 +5 +6 +7 +8 +9 +10

Promene



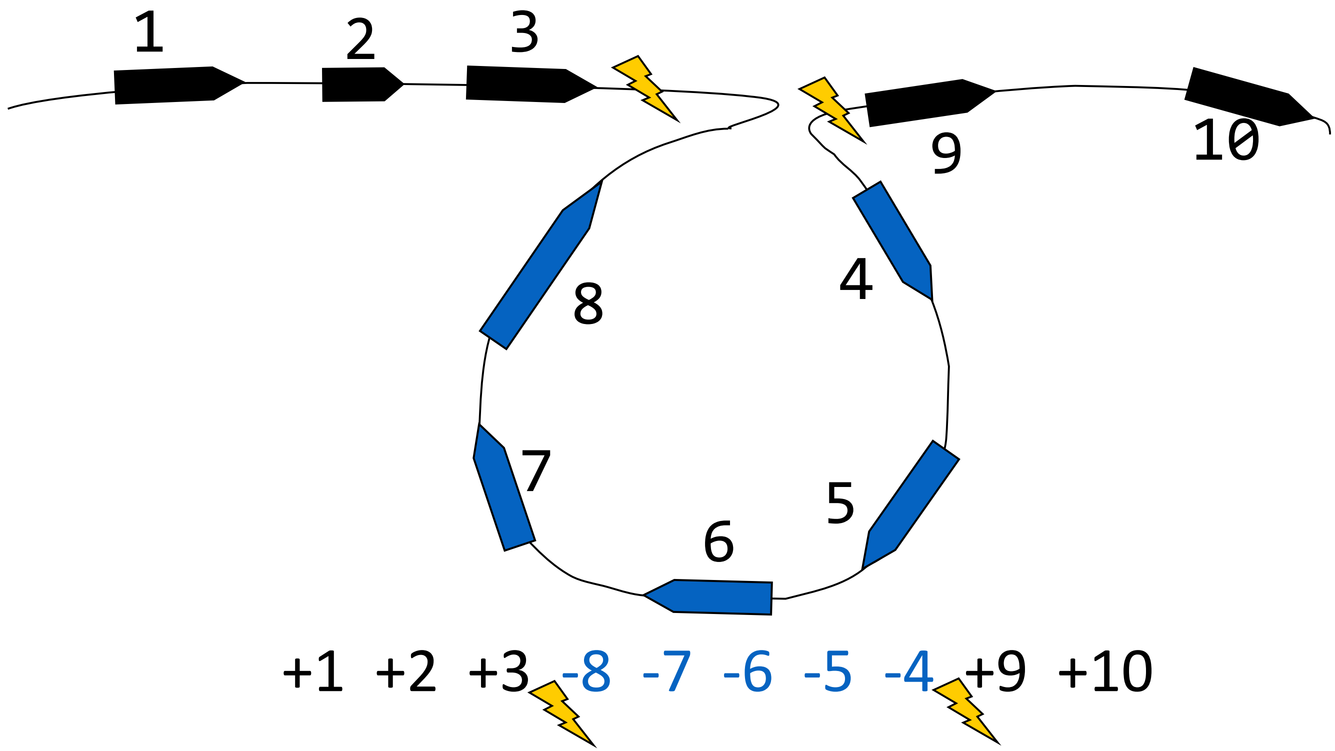
+1 +2 +3 +4 +5 +6 +7 +8 +9 +10

Promene



+1 +2 +3 -8 -7 -6 -5 -4 +9 +10

Promene



Promene u genomu su dovele do stvaranja dve *tačke prekida* ( ) koje predstavljaju poremećaj u redosledu gena u genomu.

Scenario preuređivanja sa 5 promena

Step 0:	2	-4	-3	5	-8	-7	-6	1
Step 1:	2	3	4	5	-8	-7	-6	1
Step 2:	2	3	4	5	6	7	8	1
Step 3:	2	3	4	5	6	7	8	-1
Step 4:	-8	-7	-6	-5	-4	-3	-2	-1
Step 5:	1	2	3	4	5	6	7	8

Scenario preuređivanja sa 4 promene

Step 0:	2 -4 -3 5 -8 -7 -6 1
Step 1:	2 3 4 5 -8 -7 -6 1
Step 2:	-5 -4 -3 -2 -8 -7 -6 1
Step 3:	-5 -4 -3 -2 -1 6 7 8
Step 4:	1 2 3 4 5 6 7 8

Rastojanje permutacija: najmanji broj promena potrebnih za transformisanje jedne permutacije u drugu.

Sortiranje po promenama

Step 0:	2	-4	-3	5	-8	-7	-6	1
Step 1:	2	3	4	5	-8	-7	-6	1
Step 2:	-5	-4	-3	-2	-8	-7	-6	1
Step 3:	-5	-4	-3	-2	-1	6	7	8
Step 4:	1	2	3	4	5	6	7	8

Problem sortiranja po promenama: Izračunati rastojanje između date permutacije i identične permutacije (+1 +2 … +*n*).

- **Ulaz**: Permutacija *P*.
- **Izlaz**: Rastojanje između permutacije *P* i identične permutacije.

Pohlepno sortiranje po promenama

(+1 **−7 +6 −10 +9 −8 +2** −11 −3 +5 +4)

(+1 **−2** +8 −9 +10 −6 +7 −11 −3 +5 +4)

(+1 +2 **+8 −9 +10 −6 +7 −11 −3** +5 +4)

(+1 +2 +3 **+11 −7 +6 −10 +9 −8 +5 +4**)

(+1 +2 +3 **−4** −5 +8 −9 +10 −6 +7 −11)

(+1 +2 +3 +4 **−5** +8 −9 +10 −6 +7 −11)

(+1 +2 +3 +4 +5 **+8 −9 +10 −6** +7 −11)

(+1 +2 +3 +4 +5 +6 **−10 +9 −8 +7** −11)

(+1 +2 +3 +4 +5 +6 **−7** +8 −9 +10 −11)

(+1 +2 +3 +4 +5 +6 +7 +8 **−9** +10 −11)

(+1 +2 +3 +4 +5 +6 +7 +8 +9 +10 **−11**)

(+1 +2 +3 +4 +5 +6 +7 +8 +9 +10 +11)

Pohlepno sortiranje po promenama

- Posmatrajmo sledeći primer:

(-6 +1 +2 +3 +4 +5)

(-1 +6 +2 +3 +4 +5) step 1

(+1 +6 +2 +3 +4 +5) step 2

(+1 -2 -6 +3 +4 +5) step 3

(+1 +2 -6 +3 +4 +5) step 4

…

(+1 +2 +3 +4 + -6 +5) step 9

(+1 +2 +3 +4 + -5 +6) step 10

(+1 +2 +3 +4 + +5 +6)

- Kraći način:

(-6 +1 +2 +3 +4 +5)

(-5 -4 -3 -2 -1 +6)

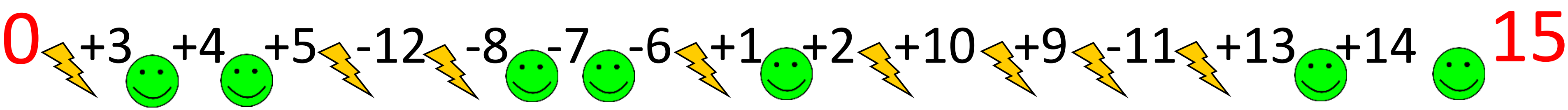
(+1 +2 +3 +4 +5 +6)

- Pohlepno sortiranje je loša aproksimacija rastojanja između dve permutacije!

Pregled

- Transformacija čoveka u miša
- Sortiranje po promenama
- **Teorema o prekidnoj tački**
- Preuređivanje u multihromozomalnim genomima
- Problem rastojanja 2-prekida
- Grafovi prekidnih tačaka
- Teorema o rastojanju 2-prekida

Susedi i prekidi



dobar redosled

loš redosled

dobar redosled

Iako su blokovi invertovani

$$adjacencies(P) + breakpoints(P) = |P| + 1$$

Koliko prekidnih tačaka ima u identičkoj permutaciji (+1 +2 … +*n*)?

Sortiranje po promenama
kao eliminacija prekidnih tačaka

									breakpoints(P)
Step 0:	2	-4	-3	5	-8	-7	-6	1	**6**
Step 1:	2	3	4	5	-8	-7	-6	1	**4**
Step 2:	-5	-4	-3	-2	-8	-7	-6	1	**4**
Step 3:	-5	-4	-3	-2	-1	6	7	8	**2**
Step 4:	1	2	3	4	5	6	7	8	**0**

Koliko prekidnih tačaka može biti eliminisano u jednoj promeni?

Eliminacija prekidnih tačaka

- Razmotrimo koliki je *minimalan* broj tačaka prekida koje se mogu eliminisati jednom promenom odnosno jednim obrtanjem
- Na primer, u permutaciji

  (0 +3 +4 +5 -12 -8 -7 -6 +1 +2 +10 +9 -11 +13 +14 15)

  unutar oblasti promene imamo pet prekidnih tačaka:

  (-12 -8) (-6 +1) (+2 +10) (+10 +9) (+9 -11)
- Nakon obrtanja

  (0 +3 +4 +5 +11 -9 -10 -2 -1 +6 +7 +8 +12 +13 +14 15)

  u promenjenom delu ćemo ponovo imati pet prekidnih tačaka, drugačijih nego pre obrtanja:

  (+11 -9) (-9 -10) (-10 -2) (-1 +6) (+8 +12)

Eliminacija prekidnih tačaka

- Razmotrimo koliki je *maksimalan* broj tačaka prekida koje se mogu eliminisati jednom promenom odnosno jednim obrtanjem

- S obzirom da se sve prekidne tačke unutar i izvan oblasti promene ostaju prekidne tačke i nakon promene, jedine tačke koje se mogu eliminisati obrtanjem su dve prekidne tačke na granicama oblasti promene

- Maksimalni broj permutacija koje se mogu eliminisati obrtanjem iznosi dva

Eliminacija prekidnih tačaka

- Pogledajmo primer gde se obrtanjem eliminiše jedna prekidna tačka

- Date su polazna i rezultujuća permutacija:

(+3 +4 +5 -12 -8 -7 -6 +1 +2 +10 +9 -11 +13 +14)

(+3 +4 +5 +11 -9 -10 -2 -1 +6 +7 +8 +12 +13 +14)

- Granični prekidi u prvoj promeni su (+5 −12) i (−11 +13), nakon promene postanu (+5 +11) koja predstavlja prekid i (+12 +13) koja predstavlja sused
- Time je smanjen broj prekidnih tačaka unutar cele permutacije za jedan.

Eliminacija prekidnih tačaka

- Pogledajmo primer gde se obrtanjem eliminišu dve prekidne tačke

- Date su polazna i rezultujuća permutacija:

(+2 -4 -3 +5 -8 -7 -6 +1)

(+2 +3 +4 +5 -8 -7 -6 +1)

- Granični prekidi su (+2 -4) i (-3 +5), nakon promene postanu (+2 +3) i (+4 +5) i oba para predstavljaju susede

- Time je smanjen broj prekidnih tačaka unutar cele permutacije za dva.

Eliminacija prekidnih tačaka

- Jedna promena može da eliminiše najviše dve prekidne tačke.
- Za niz promena, to znači da dve promene eliminišu najviše četiri prekidne tačke, tri eliminišu najviše šest prekidnih tačaka i tako dalje
- Na osnovu toga možemo zaključiti da je rastojanje permutacije P od jedinične permutacije najmanje polovina broja prekidnih tačaka u P

Teorema o prekidnoj tački: Rastojanje između permutacija
≥ *breakpoints(P)*/**2**

Eliminacija prekidnih tačaka

- U idealnom slučaju, svaka promena će eliminisati dve prekidne tačke i ukupan broj promena do identičke permutacije će biti jednaka #prekidi(P)/2

- Međutim, to nije uvek moguće. Videli smo primer kada je samo jedna prekidna tačka eliminisana obrtanjem a postoje i primeri gde se obrtanjem ne eliminiše nijedna prekidna tačka

Eliminacija prekidnih tačaka

- Teorema o prekidnoj tački određuje *minimalni* broj promena pri transformaciji permutacije P na jediničnu permutaciju
- Razmotrimo *maksimalan* broj promena
- Permutacija (+n +(n-1) … +1) ima maksimalan broj prekidnih tačaka (n+1) i za njeno sortiranje je potrebno n+1 promena

Sortiranje po promenama eliminacijom prekidnih tačaka

- Nema garancija da će svaka promena eliminisati dve prekidne tačke (step 2)
- Najveći broj promena bi bio za permutaciju (+n +(n-1) … +1) i iznosi n+1 (gornja granica)
- Donja granica: (n+1)/2
- Velika razlika između gornje i donje granice nam daje lošu aproksimaciju rastojanja polazna permutacije i identičke permutacije
- Prelazimo na drugi način za rešavanje ovog problema

Pregled

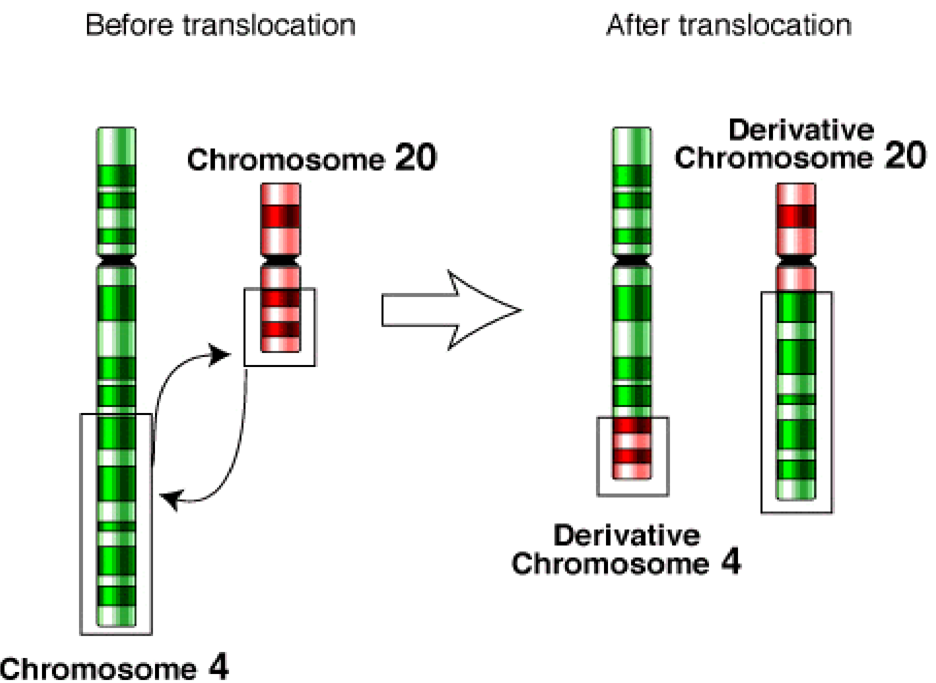
- Transformacija čoveka u miša
- Sortiranje po promenama
- Teorema o prekidnoj tački
- **Preuređivanje u multihromozomalnim genomima**
- Problem rastojanja 2-prekida
- Grafovi prekidnih tačaka
- Teorema o rastojanju 2-prekida

Preuređivanje u miltihromozomalnim genomima

- Umesto što posmatramo preuređivanje gena u okviru jednog hromozoma (hromozom X kod čoveka i miša), generalizujemo problem i posmatramo sve genome hromozoma
- U ovoj generalizaciji će biti više oblika preuređivanja blokova u genomu (do sada su bila samo obrtanja)
- Problem je naizgled komplikovaniji, u nastavku će se ispostaviti da nije tako

Preuređivanje u miltihromozomalnim genomima

translokacije

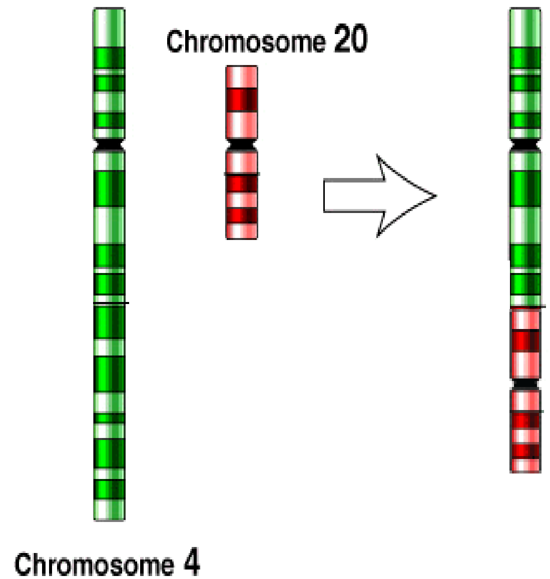


(+5 +9 +4 +8)
(−6 −1 +7 −2)

→

(+5 +9 +7 −2)
(−6 −1 +4 +8)

fuzije i fizije



(1 2)
(3 4 5)

↔

(1 2 3 4 5)

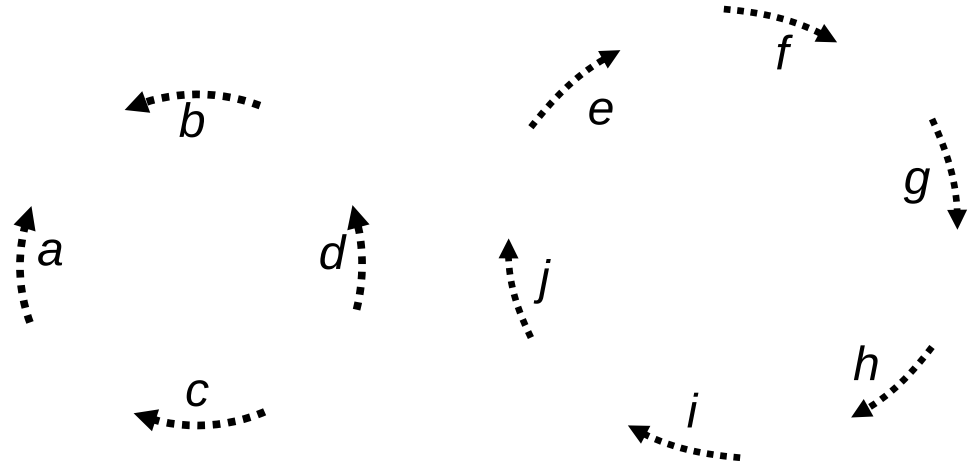
Pregled

- Transformacija čoveka u miša
- Sortiranje po promenama
- Teorema o prekidnoj tački
- Preuređivanje u multihromozomalnim genomima
- **Problem rastojanja 2-prekida**
- Grafovi prekidnih tačaka
- Teorema o rastojanju 2-prekida

Od linearnih do cirkularnih hromozoma

$(+a\ -b\ -d\ +c)$ $\quad$ $(+e\ +f\ +g\ +h\ +i\ +j)$

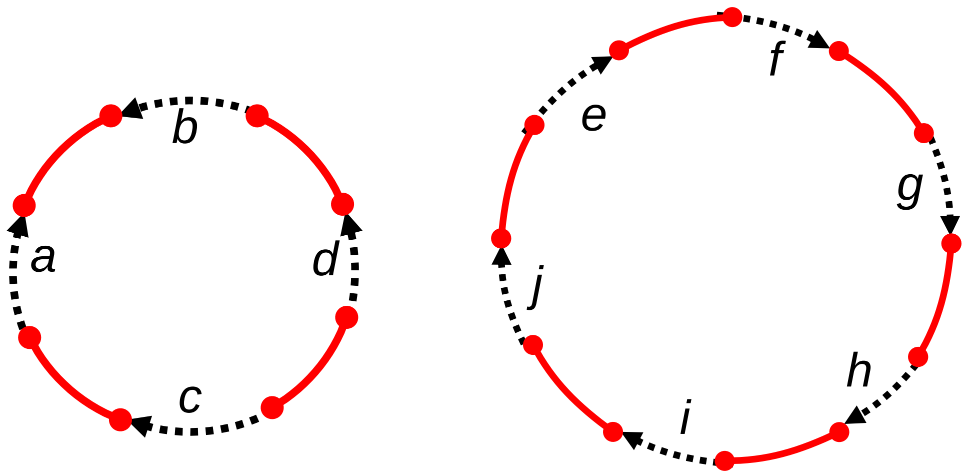
Od linearnih do cirkularnih hromozoma



$(+a\ -b\ -d\ +c)$ $(+e\ +f\ +g\ +h\ +i\ +j)$

Crne usmerene grane predstavljaju blokove sintenije.

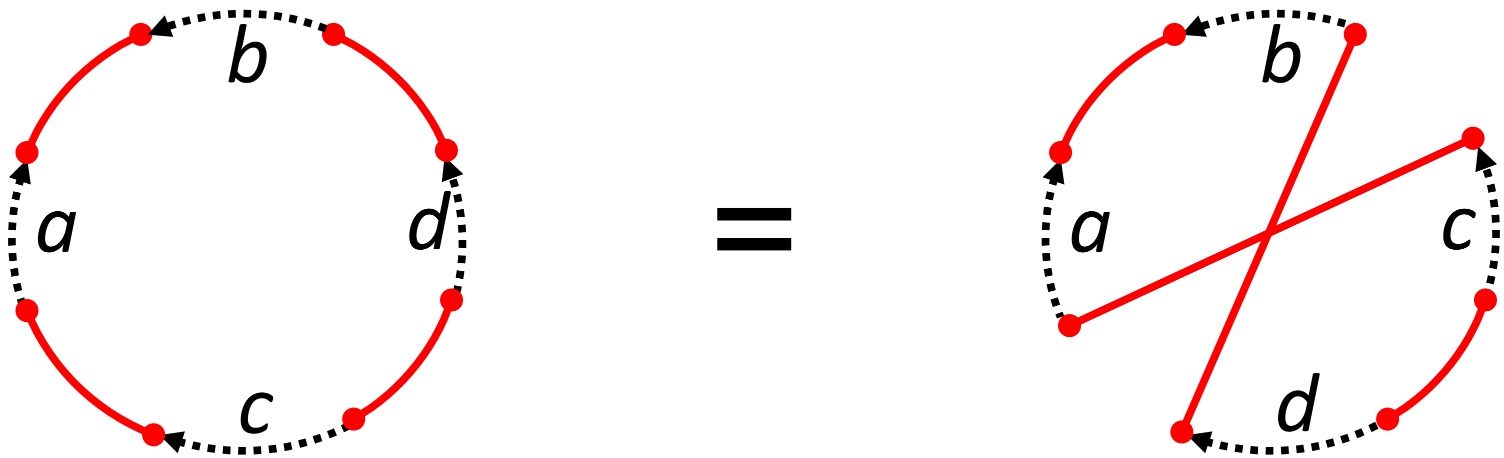
Od linearnih do cirkularnih hromozoma



$(+a\ -b\ -d\ +c)$ $(+e\ +f\ +g\ +h\ +i\ +j)$

Crne usmerene grane predstavljaju blokove sintenije.
Crvene neusmerene grane povezuju susedne blokove sintenije.

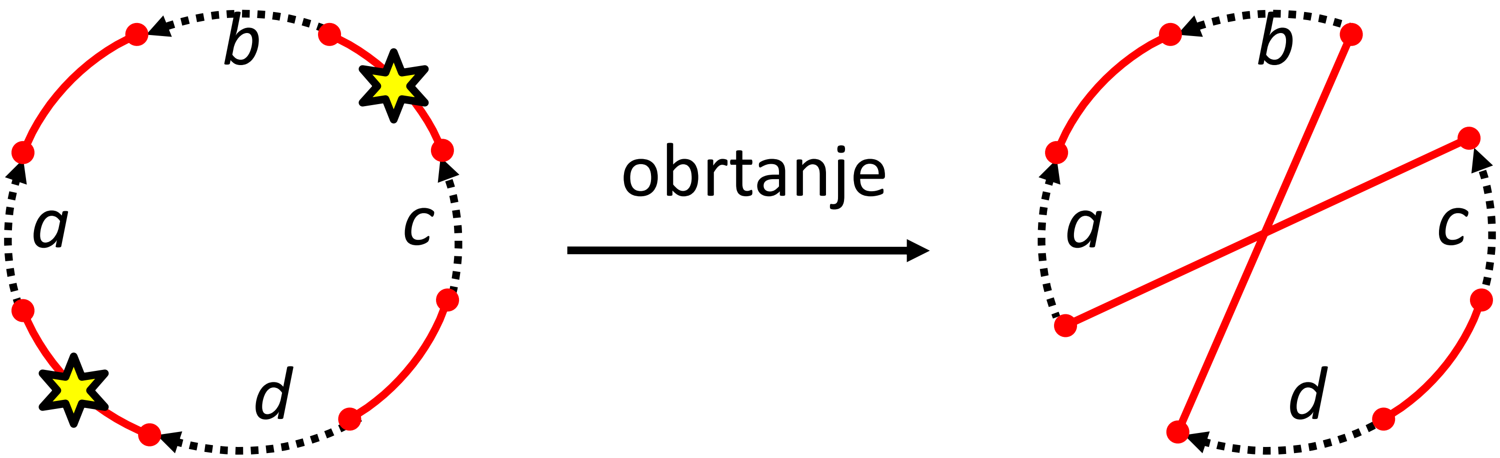
Ekvivalentne reprezentacije cirkularnog hromozoma



Q=(+a-b-d+c)

Q=(+a-b-d+c)

Obrtanje kod cirkularnog hromozoma

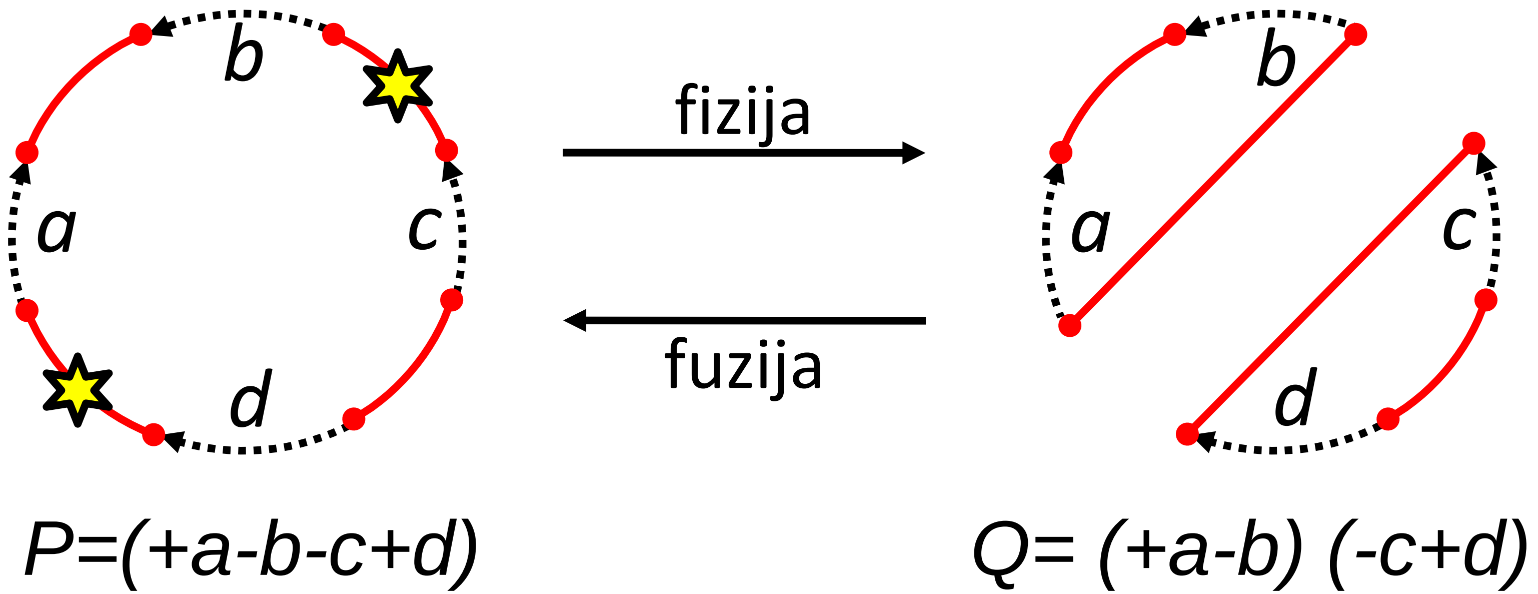


$P=(+a-b\underline{-c+d})$

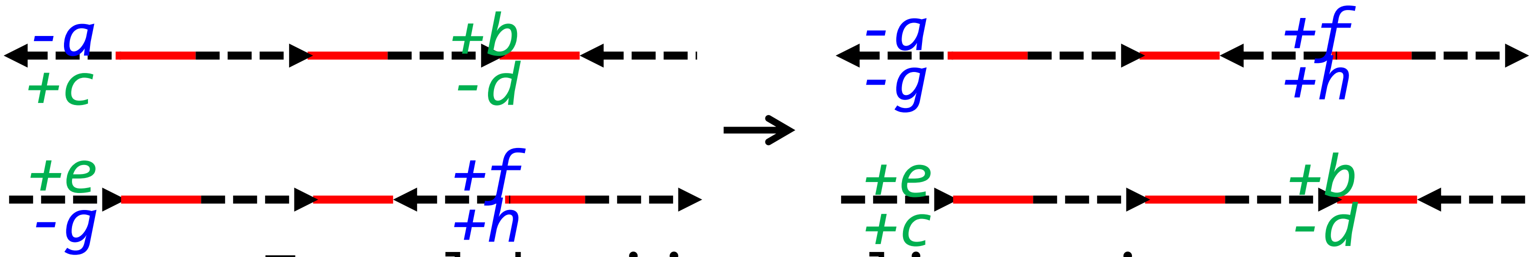
$Q=(+a-b-d+c)$

Obrtanje (-c+d u -d+c) briše dve crvene grane i zamenjuje ih sa druge dve crvene grane (čvorovi ostaju isti)

Fizija/fuzija u cirkularnom hromozomu kao zamena dve crvene grane

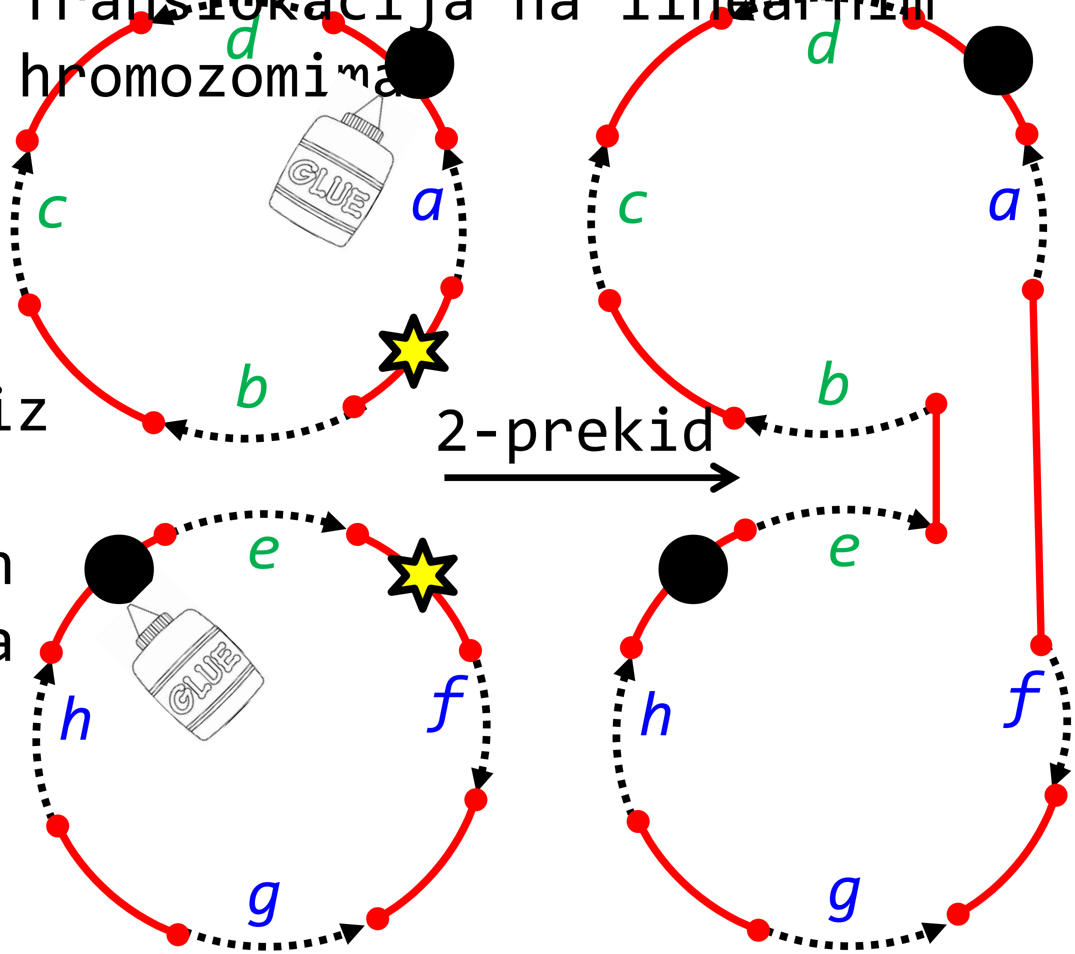


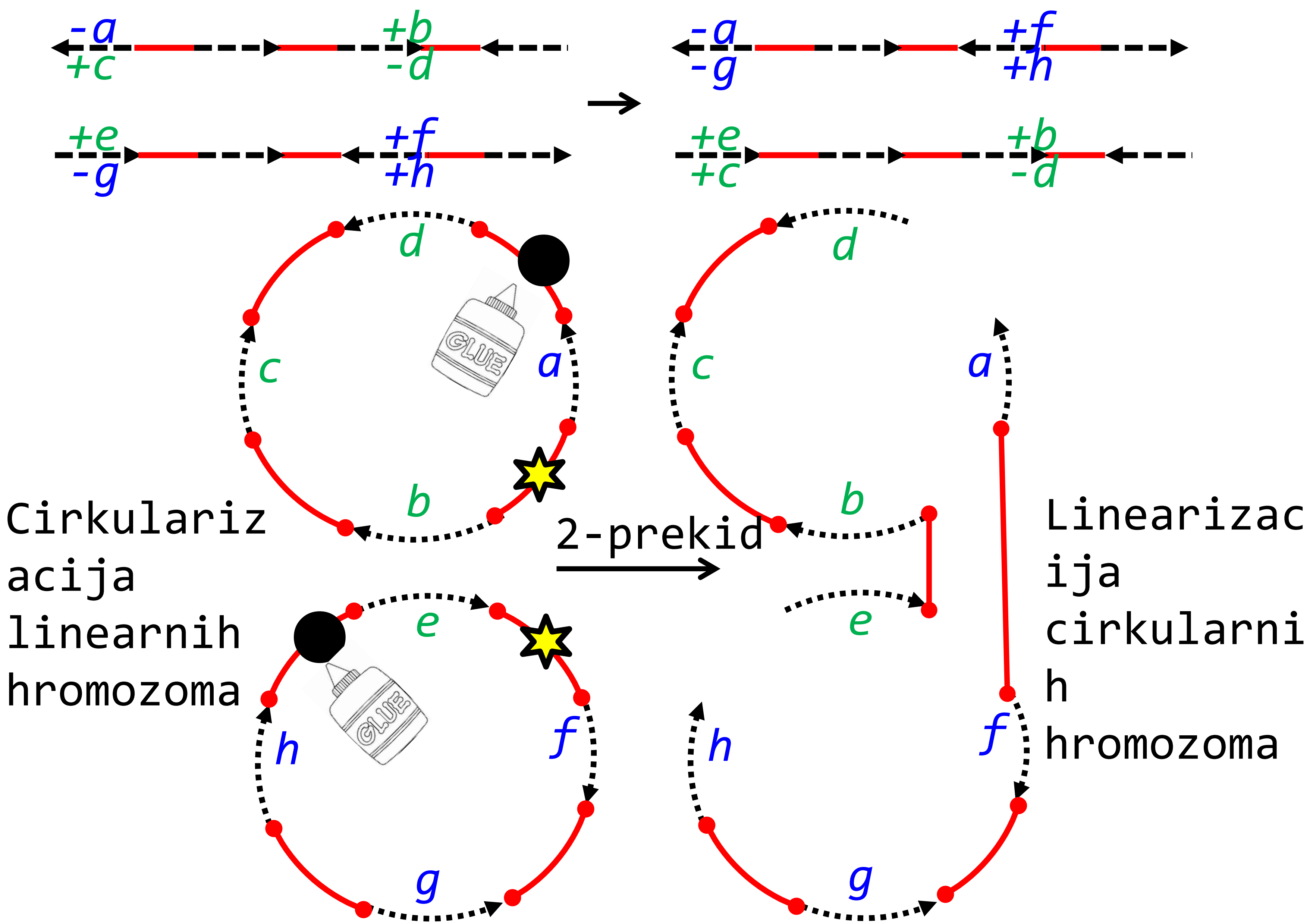
2-prekid: preuređenje koje zamenjuje dve crvene grane sa dve nove crvene grane (čvorovi ostaju isti)



Translokacija na linearnim hromozomima

Cirkularizacija linearnih hromozoma



-a
+c
+b
-d
+e
-g
+f
+h
-a
-g
+f
+h
+e
+c
+b
-d
d
c
a
b
e
h
f
g
d
c
a
b
e
h
f
g
Cirkularizacija linearnih hromozoma
2-prekid
Linearizacija cirkularnih hromozoma

- Sva četiri tipa preuređenja genoma se mogu posmatrati kao uklanjanje 2 crvene grane grafa genoma i uvodenje dve nove crvene grane na ista 4 čvora
- Operaciju zamene dve crvene grane drugim dvema crvenim granama na grafu genoma zovemo *2-prekid*
- Jedan način za opisivanje rastojanja između hromozoma *P* i *Q* je *rastojanje 2-prekida*

Rastojanje 2-prekida

Rastojanje 2-prekida *d(P,Q)*:
Minimalni broj 2-prekida koji transformišu genom *P* u genom *Q*

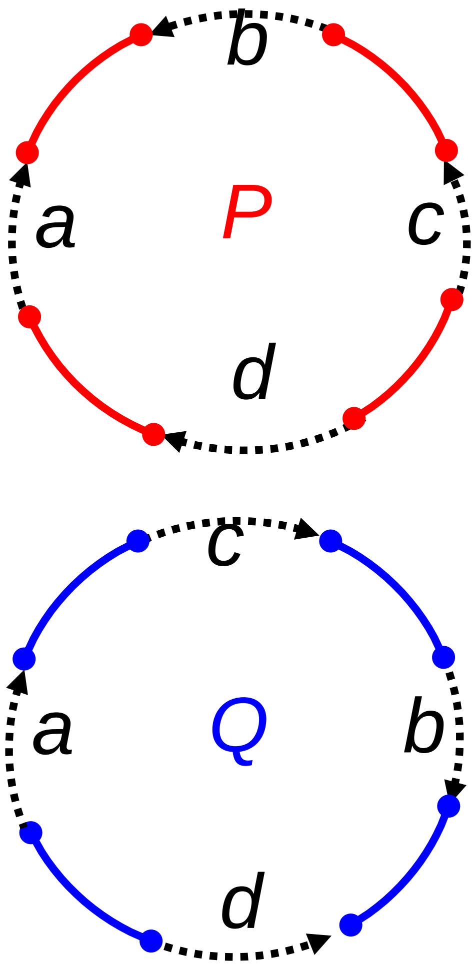
Problem rastojanja 2-prekida. Naći rastojanje 2-prekida između dva genoma.

- **Ulaz.** Dva genoma nad istim skupom blokova sintenije.
- **Izlaz.** Rastojanje 2-prekida između ovih genoma.

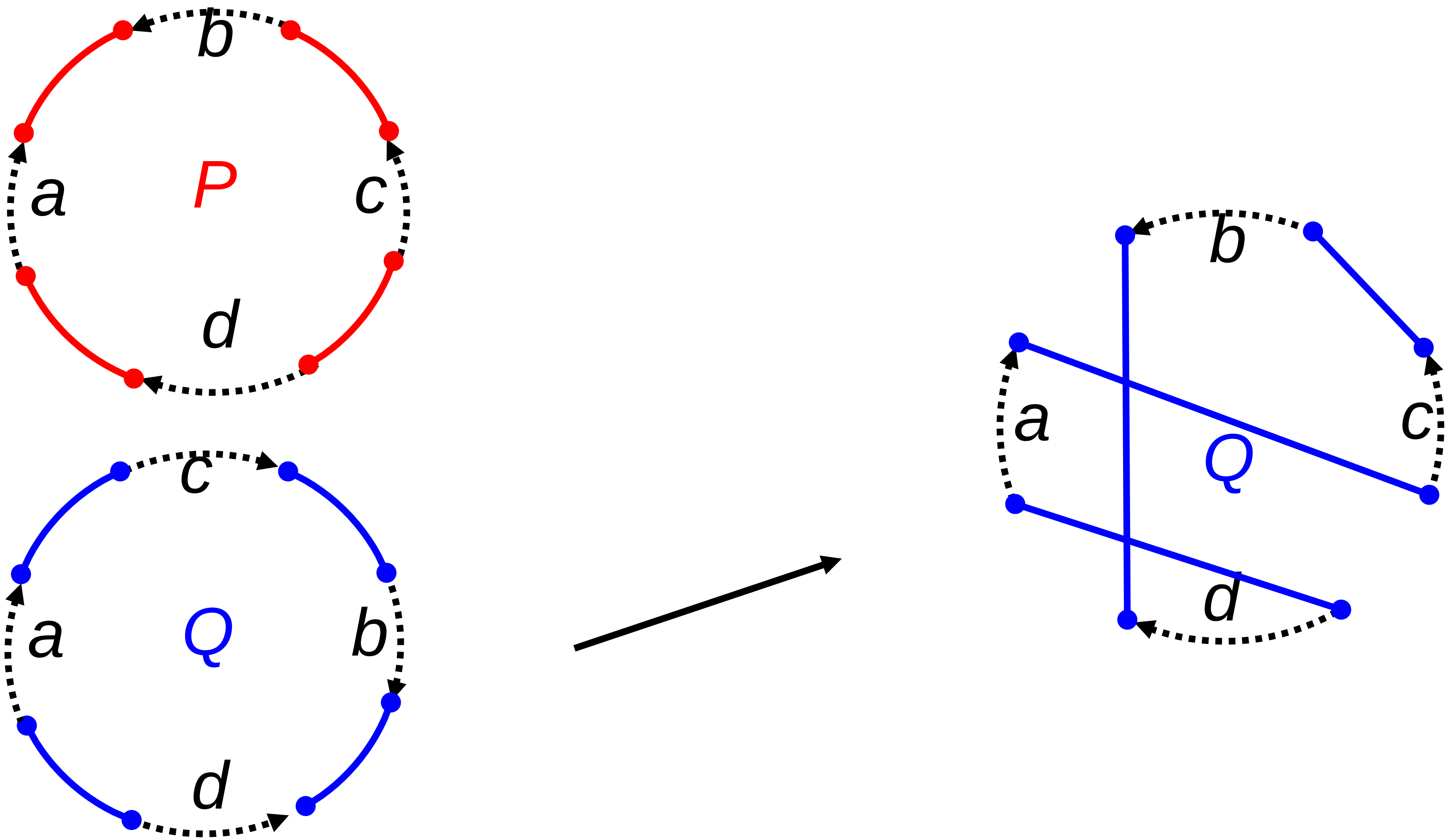
Pregled

- Transformacija čoveka u miša
- Sortiranje po promenama
- Teorema o prekidnoj tački
- Preuređivanje u multihromozomalnim genomima
- Problem rastojanja 2-prekida
- **Grafovi prekidnih tačaka**
- Teorema o rastojanju 2-prekida

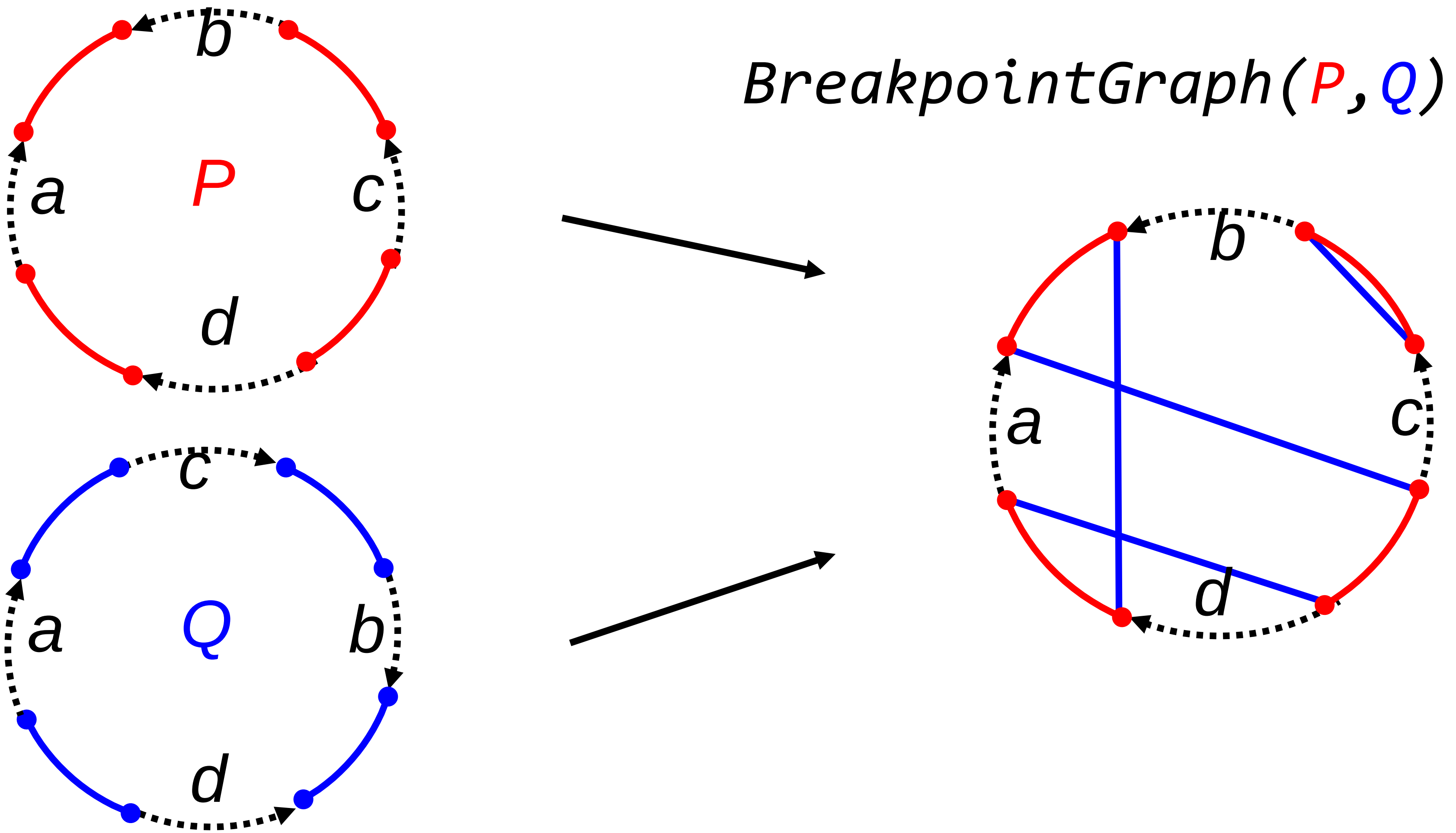
Poređenje genoma *P* i *Q*



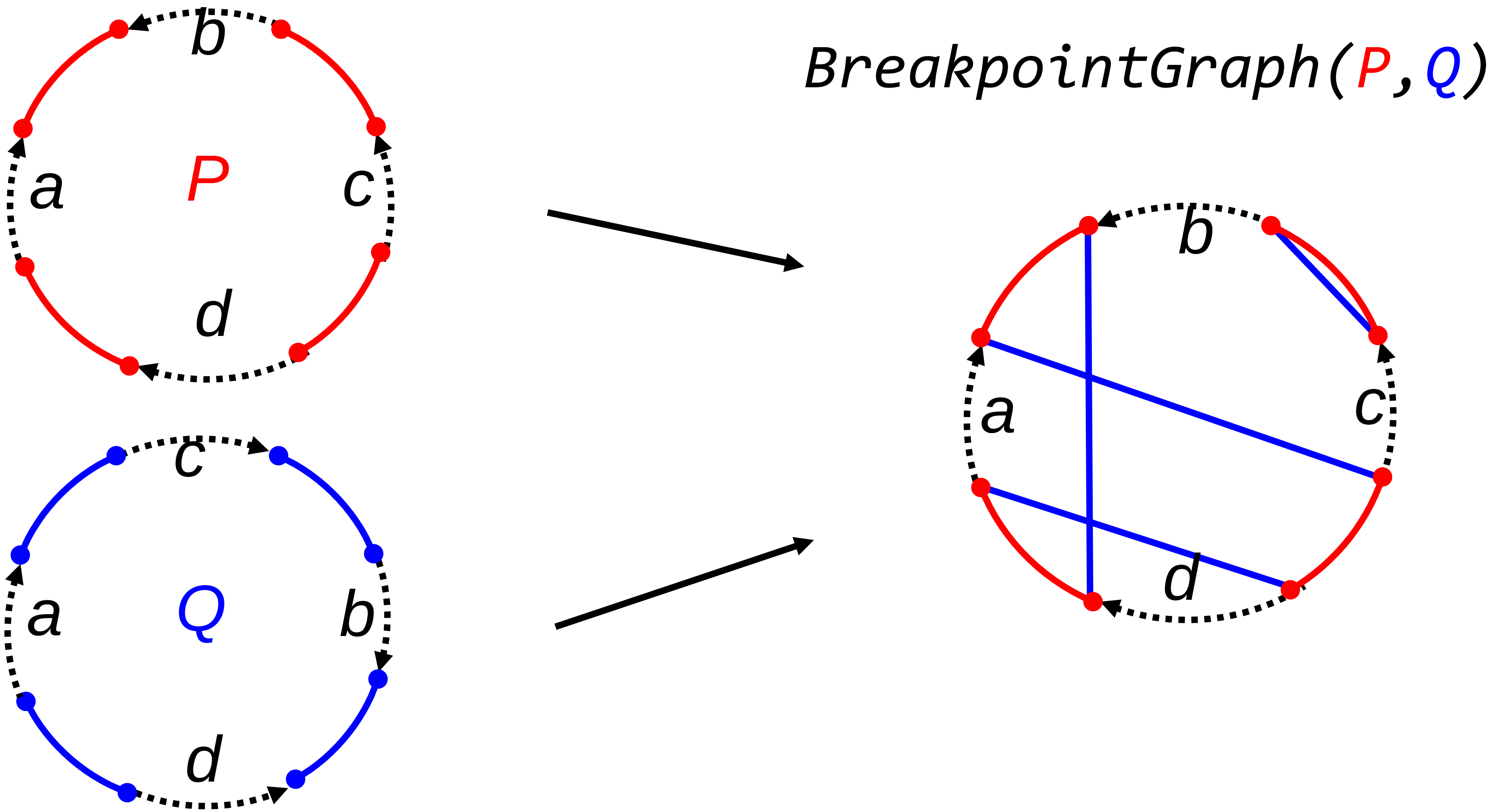
Drugačija reprezentacija Q



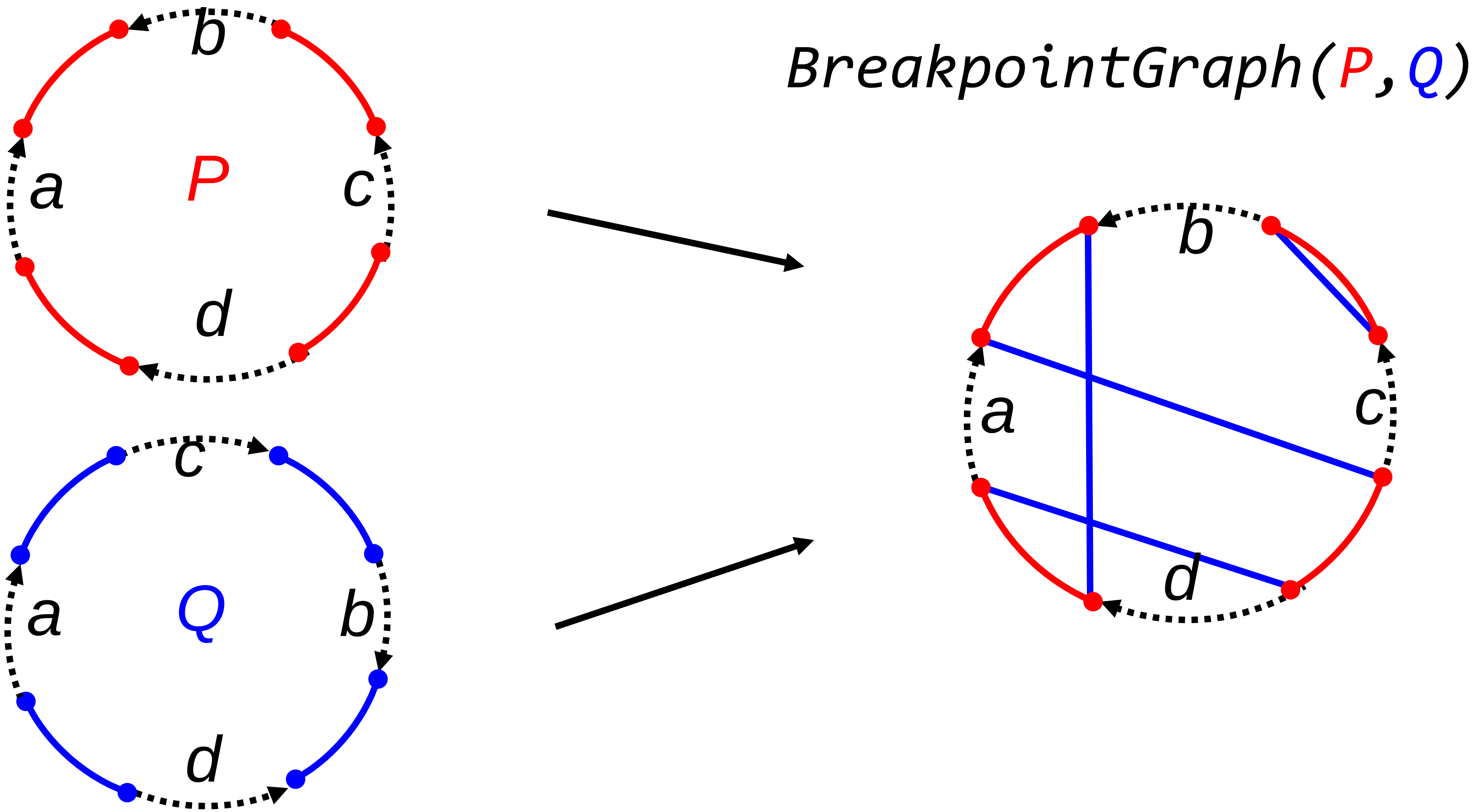
Nadgradnja *P* i *Q*



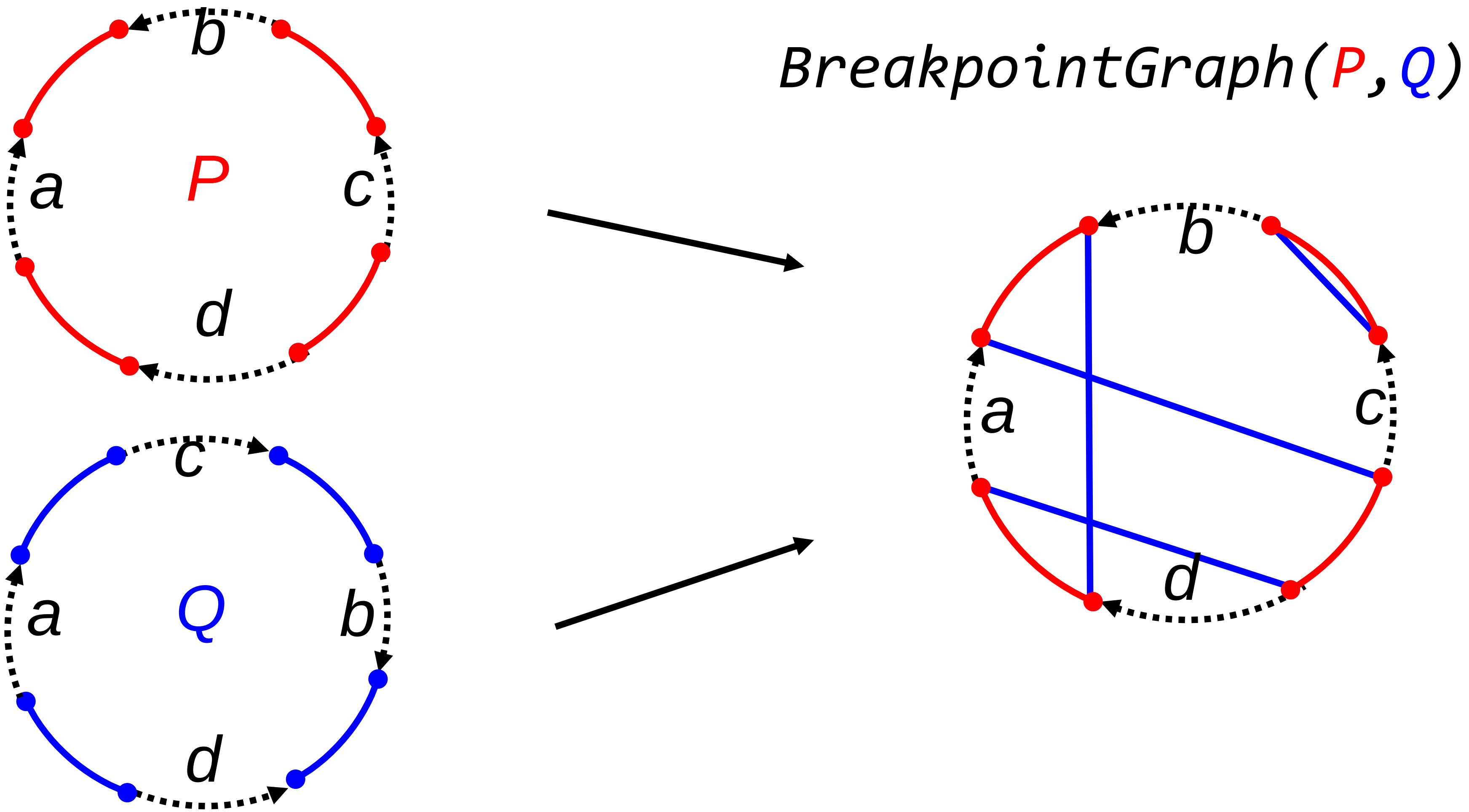
Crvene i crne grane u *breakpoint* grafu formiraju genom *P*



Plave i crne grane u *breakpoint* grafu formiraju genom *Q*

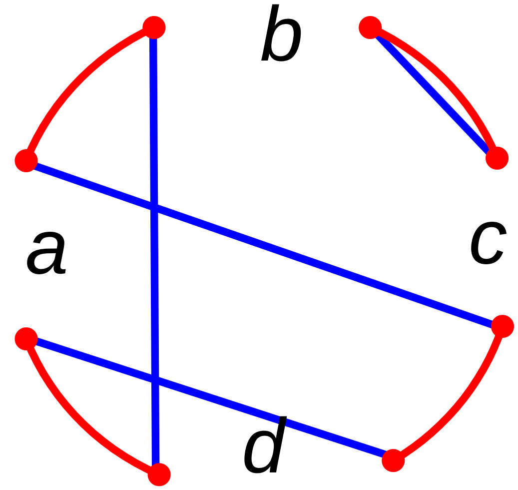


A crvene i plave grane?



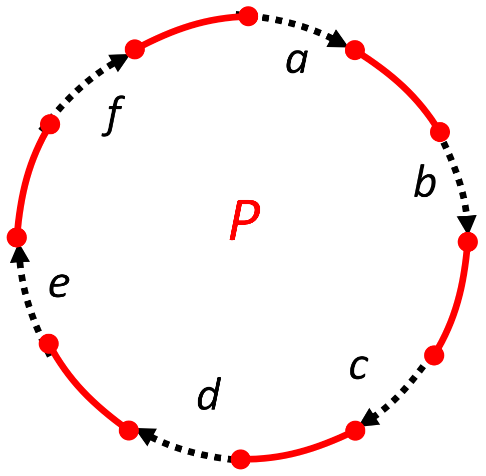
Alternirajući crveno-plavi ciklusi

BreakpointGraph(P,Q)

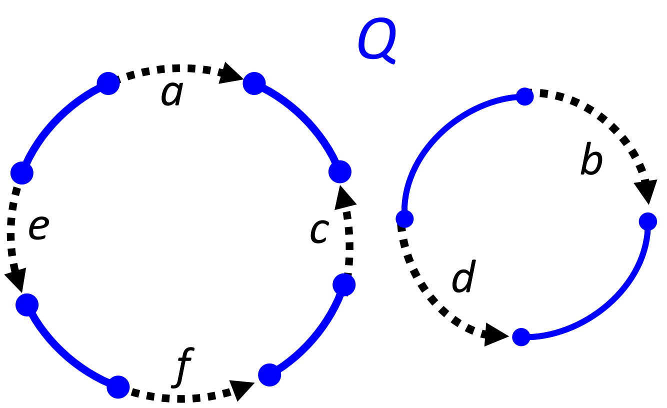


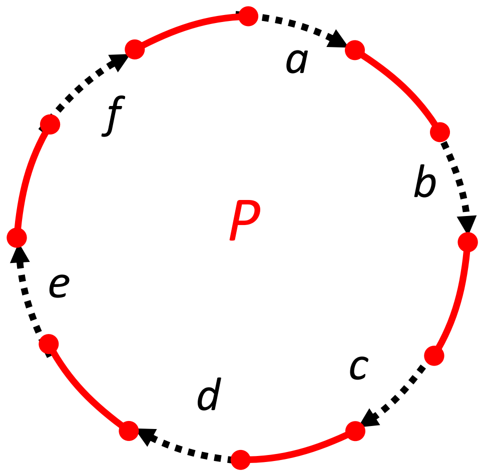
Crvene i plave grane formiraju **alternirajuće** crveno-plave cikluse.

cycle(P,Q): broj alternirajućih crveno-plavih ciklusa.

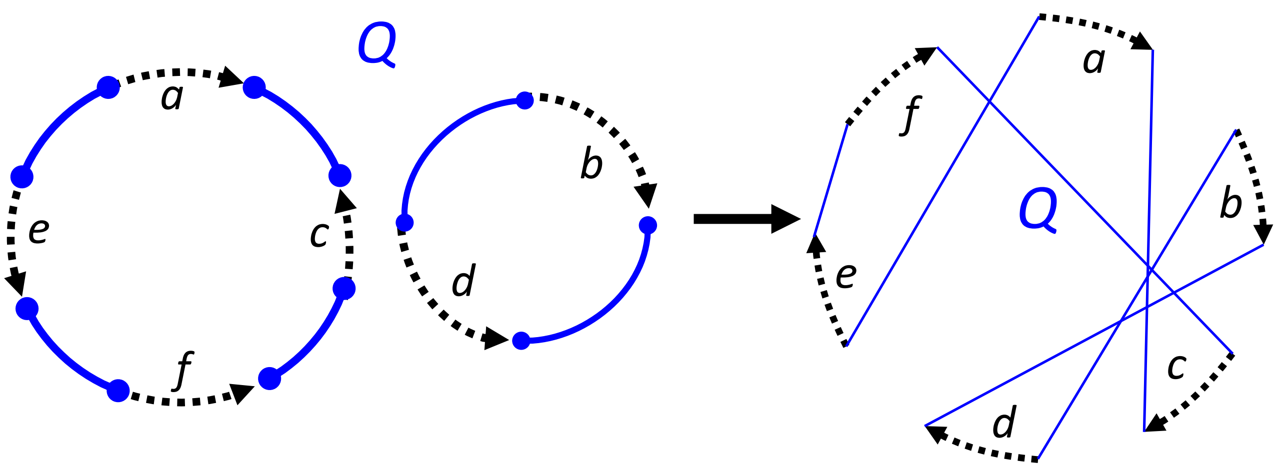


Šta predstavlja *cycle*(*P*,*Q*)?

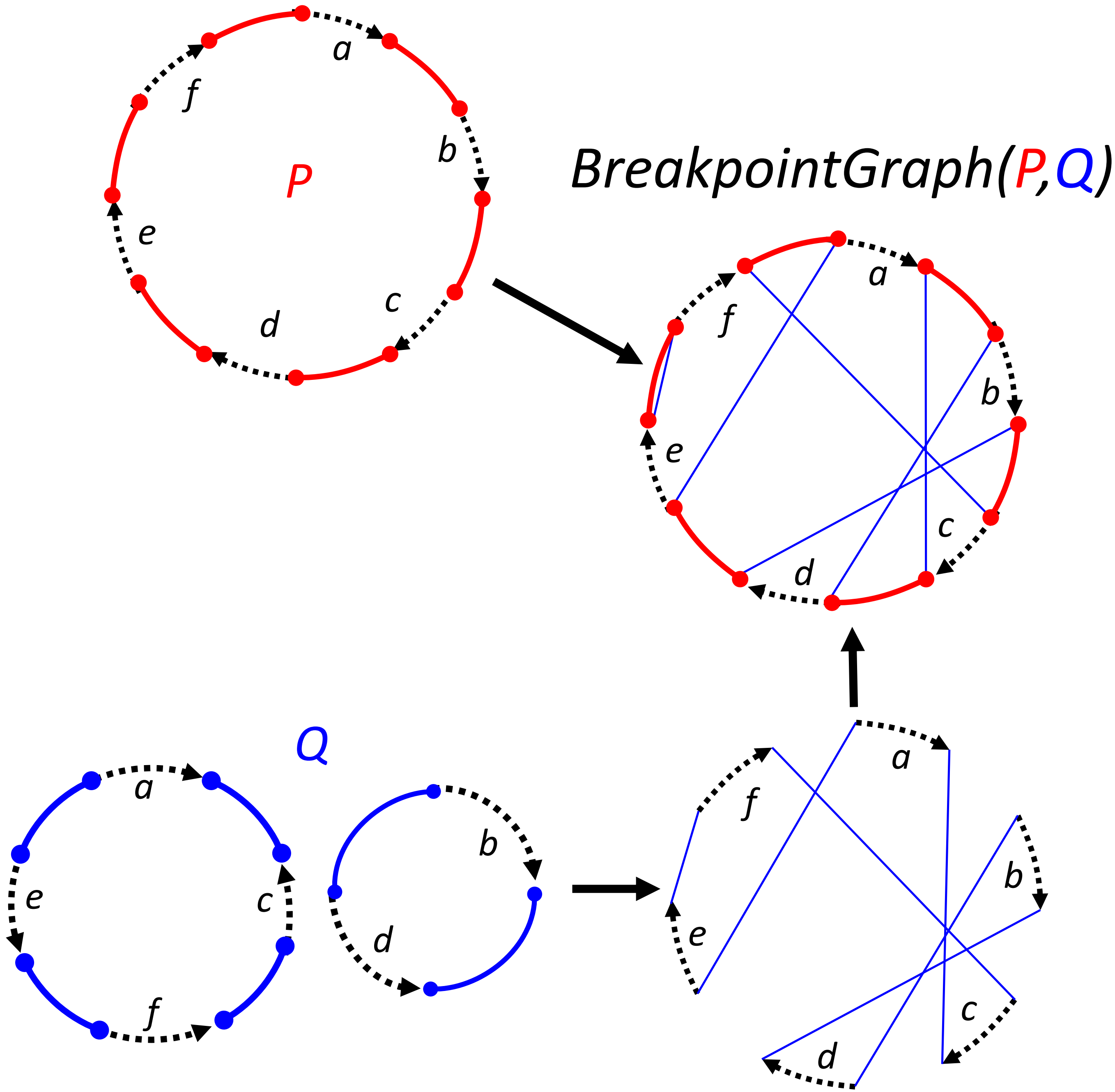


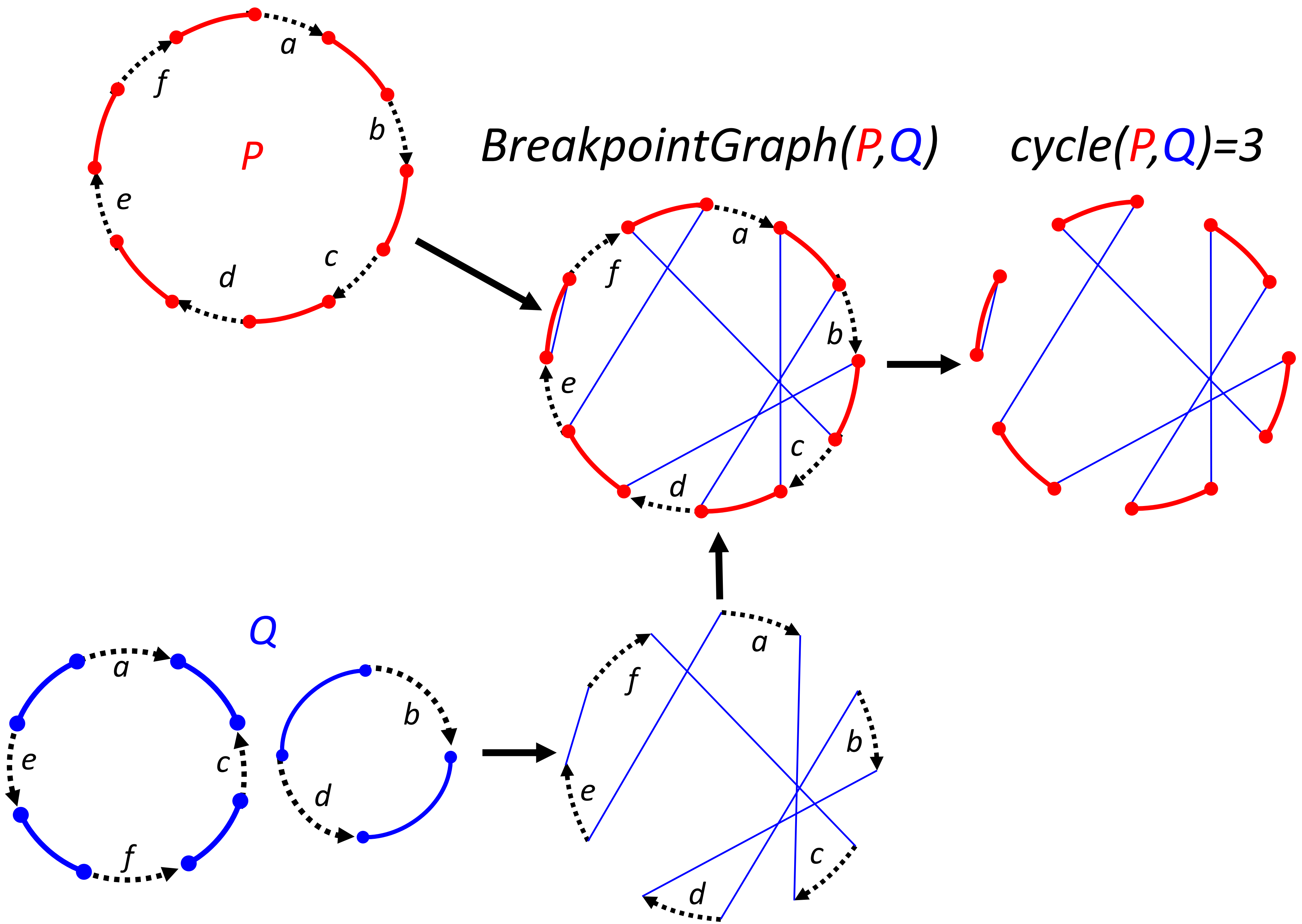


Poređamo crne grane genoma *Q*
u isti redosled kao u genomu *P*



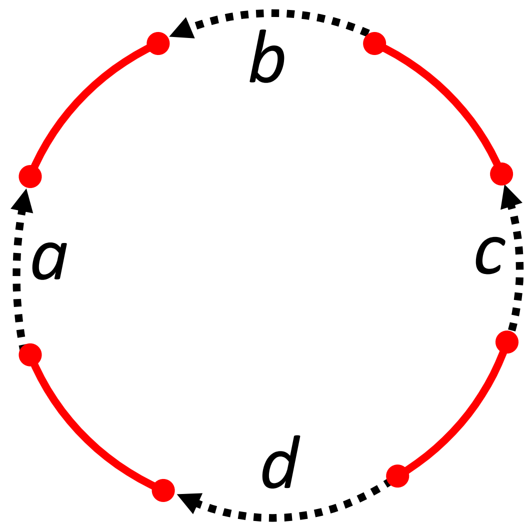
BreakpointGraph(P,Q)





P
BreakpointGraph(P,Q)
cycle(P,Q)=3
Q
a
b
c
d
e
f

Za dato *P*, koje *Q* maksimizuje *cycle*(*P*,*Q*)?

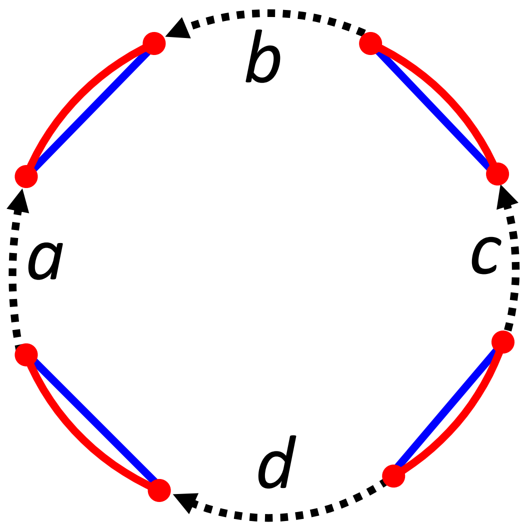


P=(+a-b-c+d)

Za dato *P*, koje *Q* maksimizuje *cycle(P,Q)*?

BreakpointGraph(P,P)

Identički *breakpoint* graf



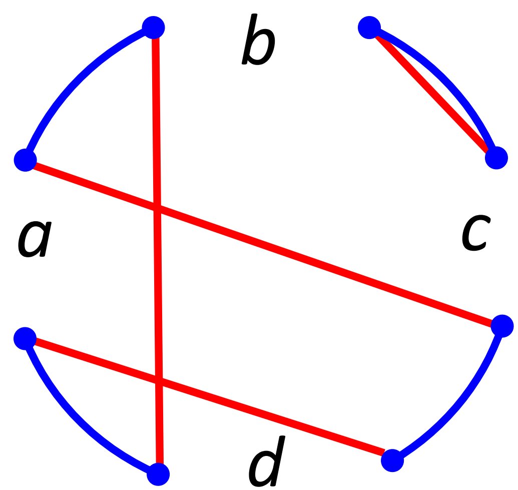
cycle(P,P)=
#blocks=4

P=(+a-b-c+d)

Preuređenje genoma utiče na crveno-plave cikluse

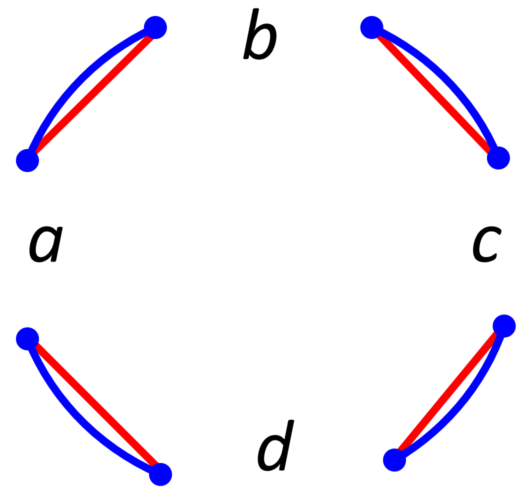
Svaka transformacija *P* → *Q* odgovara transformaciji :

BreakpointGraph(P,Q)



Niz 2-prekida koji transformišu
P u *Q*

BreakpointGraph(Q,Q)



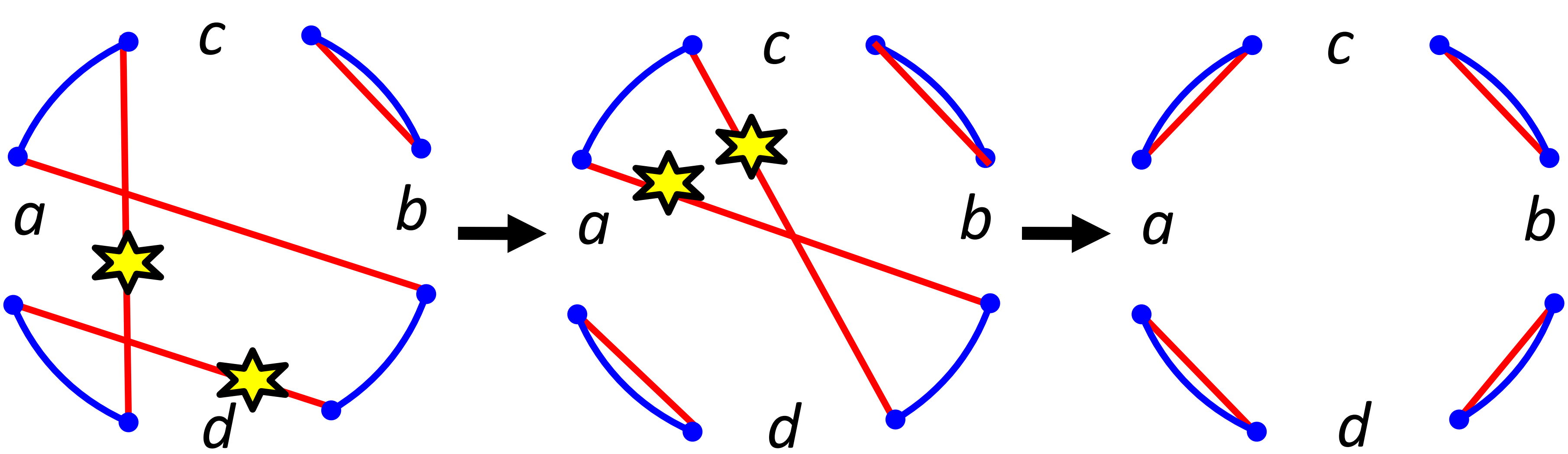
cycle(P,Q)=2 → *cycle(Q,Q)= 4*

Preuređenja menjaju *cycle(P,Q)*

P=(+a –b –c ***+d****)* → *P'=(+a* ***–b –c*** *-d)* → *P''=Q=(+a +c +b -d)*

BreakpointGraph(P,Q) → *BreakpointGraph(P',Q)* → *BreakpointGraph(Q,Q)*

cycle(P,Q)=2 → *cycle(P',Q)=3* → *cycle(Q,Q)=4*



Pregled

- Transformacija čoveka u miša
- Sortiranje po promenama
- Teorema o prekidnoj tački
- Preuređivanje u multihromozomalnim genomima
- Problem rastojanja 2-prekida
- Grafovi prekidnih tačaka
- **Teorema o rastojanju 2-prekida**

Sortiranje po 2-prekidima

2-prekidi

$P \rightarrow \ldots \rightarrow Q$

BreakpointGraph(P,Q) → ... → *BreakpointGraph(Q,Q)*

cycle(P,Q) → ... → *cycle(Q,Q)=blocks(Q,Q)*

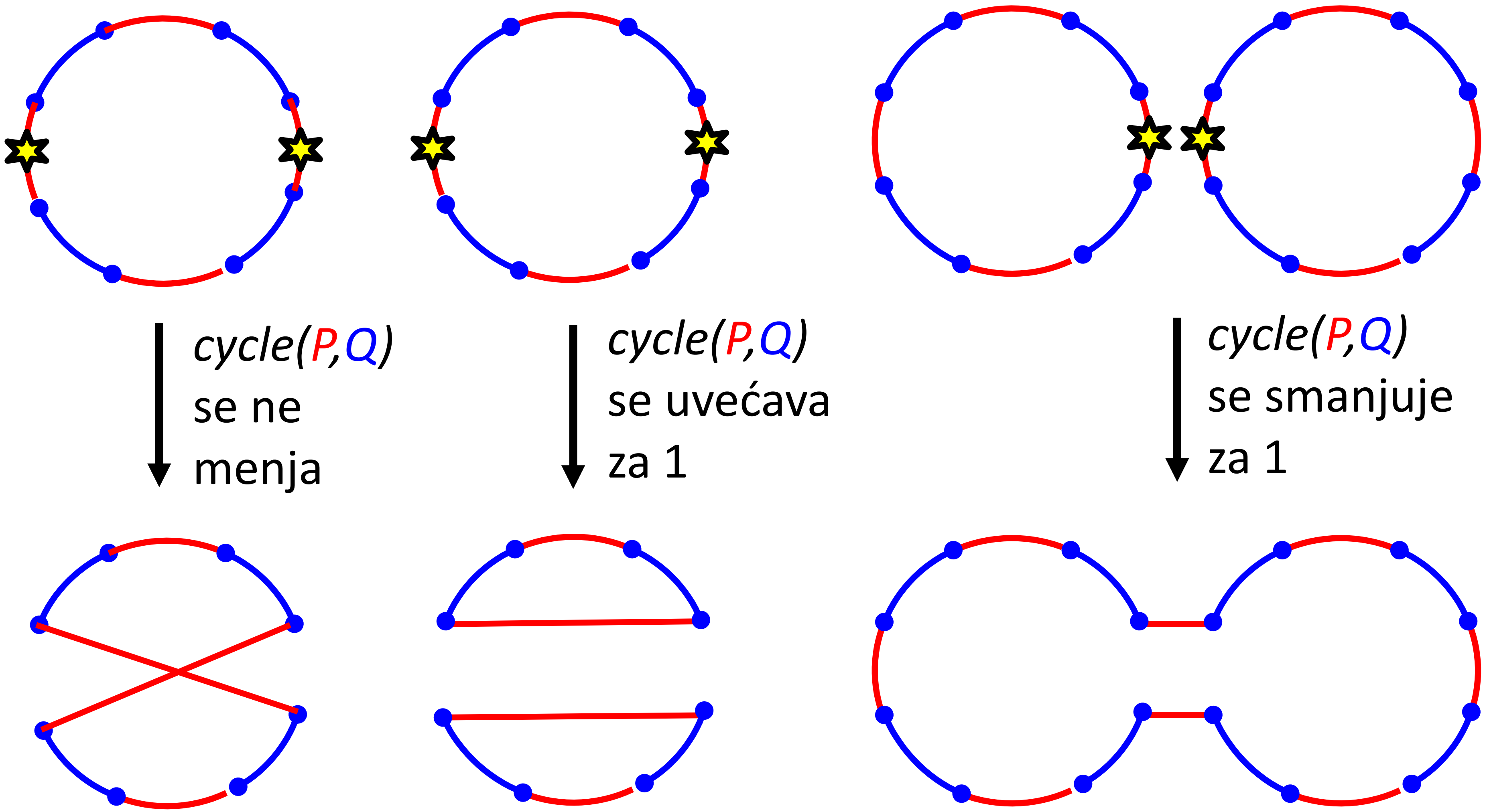
blocks(P,Q) – broj blokova koji učestvuje u izgradnji P i Q

broj `crveno-plavih` ciklusa se uvećava za

blocks(P,Q) - cycle(P,Q)

Koliko može svaki 2-prekid da doprinese ovom uvećanju?

2-prekid može izmeniti *cycle(P,Q)* za 1



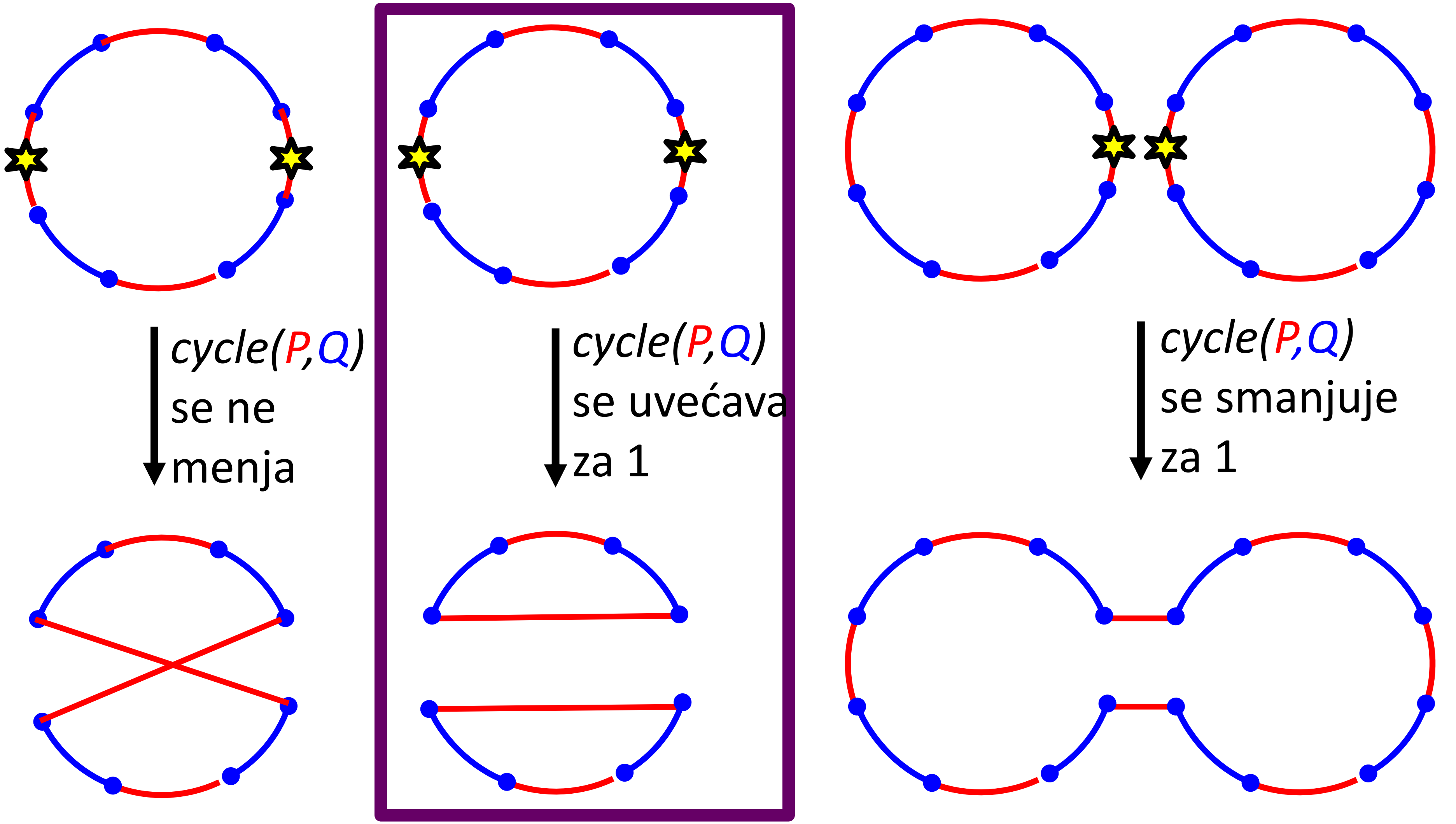
- Podsetimo se da u slučaju jednog hromozoma postoje permutacije za koje nijedno obrtanje ne umanjuje broj prekidnih tačaka i da se iz tog razloga u sortiranju po promenama javljaju koraci u kojima se broj prekidnih tačaka ne menja
- Analizirajmo po ovom pitanju generalizovani slučaj sa multihromozomalnim genima i 2-prekidima
- Znamo da svaki 2-prekid može povećati broj ciklusa za najviše 1
- Pitamo se da li je *uvek* moguće pronaći 2-prekid koji će povećati broj ciklusa za 1. Sledeća teorema daje potvrdan odgovor na ovo pitanje.

- **Teorema o rastojanju 2-prekida**:

Rastojanje 2-prekida d(P, Q) je jednako blocks(P, Q) – cycle(P, Q)

- **Dokaz**
- Svaka tranformacija genoma P u genom Q, koja se sastoji od niza pojedinačnih promena, povećava broj alternirajućih ciklusa za ukupno *blocks(P, Q) – cycle(P, Q)*
- Rastojanje 2-prekida d(P, Q) predstavlja minimalni broj 2-prekida koji genom P transformišu u genom Q, pa možemo reći da je d(P, Q) ≥ blocks(P, Q) – cycle(P, Q)
- Jednakost bi važila kada svaka promena povećava broj alternirajućih ciklusa za tačno 1, a veće bi važilo kada promena ne povećava broj ciklusa
- Razmotrimo da li *uvek* možemo da odaberemo promenu koja povećava broj ciklusa odnosno da li takva promena uvek postoji.
- Ako P nije jednako Q, to znači da u njihovom prekidnom grafu postoji bar jedan neidentički alternirajući ciklus, odnosno ciklus koji sadrži više od jedne crvene i jedne plave grane
- Svaki neidentički ciklus se u sredini, može podeliti na dva ciklusa
- To znači da dok je P različito od Q, postoji 2-prekid koji će povećati broj ciklusa za 1 => važi d(P, Q) = blocks(P, Q) – cycle(P, Q)

Postoji 2-prekid povećanje veličine *cycle(P,Q)* za 1



Rastojanje 2-prekida između genoma čoveka i miša

- Genomi čoveka i miša se mogu rastaviti na 280 blokova sintenije (dužine bar pola miliona nukleotida)
- *Breakpoint* graf nad ovim blokovima ima ukupno 35 ciklusa
- Na osnovu teoreme o rastojanju 2-prekida:

$$d(H,M) = blocks(H,M) - cycle(H,M)$$
$$= \mathbf{280-35=245}$$

- Postoje različite verzije scenarija sa 245 koraka.
- Pravi evolutivni scenario je možda imao i više od 245 koraka.

- Slajdovi pokrivaju poglavlje 6 knjige *Bioinformatics Algorithms: an Active Learning Approach*
- Sadržaj slajdova je preuzet sa zvaničnih prezentacija autora i dodatno prilagođen