

BIOINFORMATIKA

14. april 2020..

Glava 1

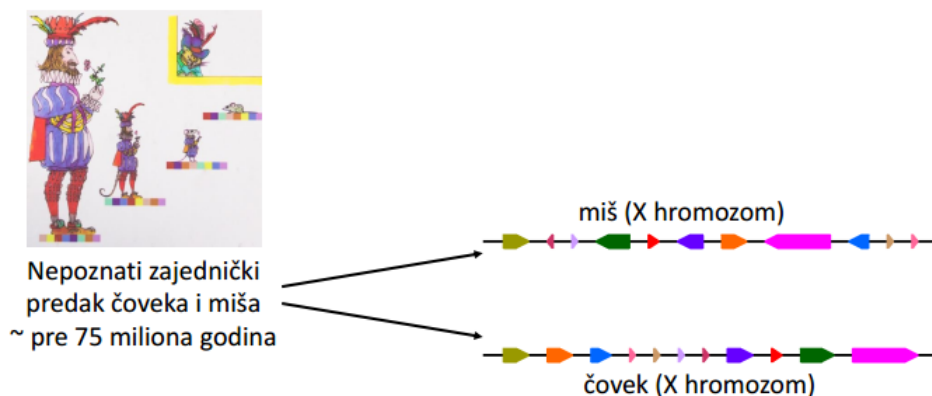
Postoje li osetljivi delovi u ljudskom genomu?

1.1 Transformacija čoveka u miša

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** (*genome rearrangements*).

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.

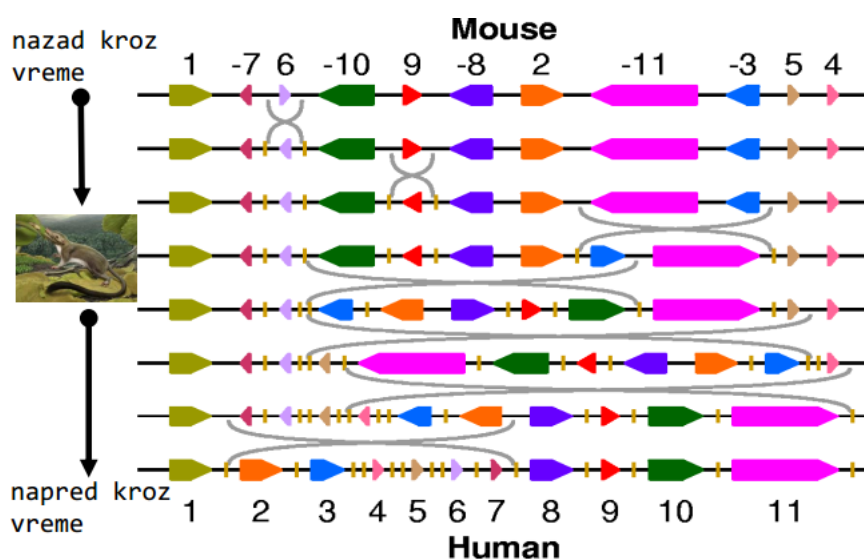
Kako bismo pojednostavili poređenje dva genoma, ograničićemo se na poređenje samo jednog hromozoma miša i čoveka, i to hromozoma X. Ovaj hromozom je jedan od dva hromozoma koji određuju pol i tokom evolucije je zadržao skoro sve gene. Slični geni, često i po nekoliko stotina njih, grupišu se često jedan pored drugog na hromozomu i zadržavaju ovaj redosled u genomima različitih vrsta. Ovakvu grupu gena nazivamo **blokom sintenije**. Deo mišjeg i ljudskog hromozoma X, sa označenim blokovima sintenije (ukupno 11), prikazan je na slici 1.1. Ovakvom reprezentacijom hromozoma sveli smo problem poređenja X hromozoma na poređenje dve sekvence sa po 11 jedinica (blokova sintenije) umesto na poređenje po nukleotidima kojih ima oko 150 miliona.



Slika 1.1: X hromozom miša i čoveka

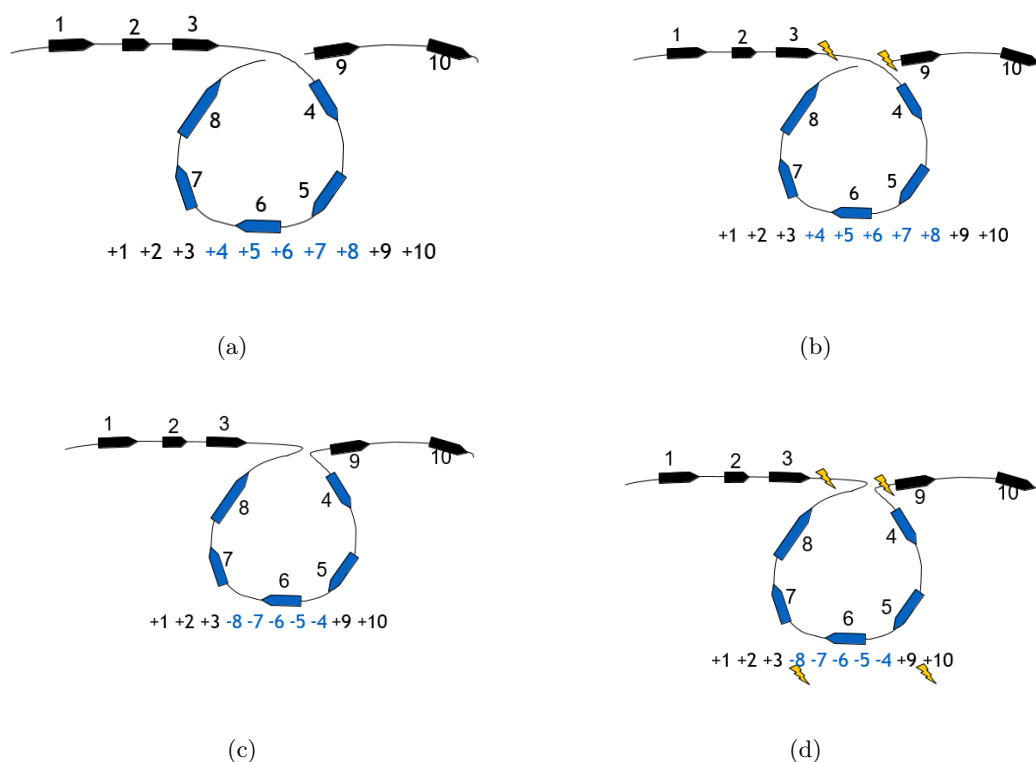
1.2 Sortiranje po promenama

Jedan scenario preuređenja genoma od mišjeg do čovekovog u 7 koraka prikazan je na slici 1.2 . Vidimo da su blokovi sintenije prikazani kao strelice što znači da imaju usmerenje i da se tokom preuređenja genoma taj smer može promeniti. Svaki blok 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.



Slika 1.2: Završena transformacija

Krećući se kroz promene od genoma miša, najpre idemo unazad kroz vreme, do zajedničkog pretka, a onda unapred, do genoma čoveka. Važno je istaći da je ovo 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, prateći princip Okamove oštrice, pretpostavićemo da je tokom evolucije bilo najmanje moguće promena između genoma miša i čoveka, odnosno između



Slika 1.3: Modelovanje promene

zajedničkog pretka i svake od ove dve vrste. Razmotrimo pitanje koji je minimalan broj koraka neophodan za transformaciju jednog genoma u drugi. Formalno, svaki genom možemo predstaviti 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)

Promenu u genomu možemo predstaviti kao na slici 1.3. Dolazi do uvrtnja sekvence od bloka +4 do bloka +8 (a), zatim se hromozom prekida na mestima uvrtnja (b), nakon toga se uspostavljaju nove veze između pokidanih delova (c) i dobija se rezultujuća sekvencu sa izmenjenim redosledom i usmerenjem blokova od +4 do +8 (d). Mesta na kojima dolazi do uspostavljanja novih veza nazivamo *tačkama prekida*.

Na slikama 1.4 i 1.5 prikazana su dva različita scenarija transformacije između genoma miša i čoveka. Postoji veliki broj mogućih scenarija transformacija, a od svih njih izdvojimo onu transformaciju koja uključuje najmanji broj promena. Broj promena koji nastaje u takvoj transformaciji zvaćemo *rastojanjem permutacija*.

Step 0:	2	<u>-4</u>	<u>-3</u>	5	-8	-7	-6	1
Step 1:	2	3	4	5	<u>-8</u>	<u>-7</u>	<u>-6</u>	1
Step 2:	2	3	4	5	6	7	8	<u>1</u>
Step 3:	2	3	4	5	6	7	8	-1
Step 4:	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
Step 5:	1	2	3	4	5	6	7	8

Slika 1.4: Scenario sa 5 promena

Step 0:	2	<u>-4</u>	<u>-3</u>	5	-8	-7	-6	1
Step 1:	2	3	4	5	-8	-7	-6	1
Step 2:	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	1
Step 3:	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>	6	7	8
Step 4:	1	2	3	4	5	6	7	8

Slika 1.5: Scenario sa 4 promene

Definicija 1.1. *Rastojanje permutacija* je najmanji broj promena potrebnih za transformisanje jedne permutacije u drugu.

Naredni problem koji posmatramo je **problem sortiranja po promenama** koji predstavlja izračunavanje rastojanja između date permutacije i identične permutacije (+1 +2 ... +n)

Ulaz: permutacija P

Izlaz: rastojanje između permutacije P i identične permutacije

Jedan pristup rešavanju ovog problema je pohlepna strategija. S obzirom da datu permutaciju treba svesti na identičnu, ovaj problem možemo posmatrati kao problem sortiranja (slika 1.6). Prvi korak je da se izvrši promena koji postavlja +1 na pravo mesto (na prvu poziciju), a zatim slede promene koji postavljaju +2 na drugu poziciju, i tako dalje. Na primer, element 1 je već na pravom mestu i ima ispravan znak (+) u X hromozomu miša, ali element 2 nije na ispravnom položaju. Element 1 možemo zadržati fiksiran i premestiti element 2 na pravi položaj jednom promenom. Još jedna promena je potrebna da bi element 2 imao ispravan znak. Daljim iteriranjem postupka dovodimo sve veće elemente na njihove ispravne pozicije.

(+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)

Slika 1.6: Pohlepno sortiranje

Definicija 1.2. Element k u permutaciji $P = (p_1, p_2, \dots, p_n)$ je **sortiran**, ako je $p_k = k$, a u suprotnom je **nesortiran**.

Definicija 1.3. Permutacija P je **k -sortirana**, ako je prvih $k-1$ elemenata sortirano, a k -ti element nesortiran.

Sledeći primer pokazuje da je pohlepno sortiranje loša aproksimacija rastojanja između dve permutacije (1.7) jer je ponekad moguće naći mnogo jednostavniji način (1.8).

(-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)

Slika 1.7: Pohlepno sortiranje

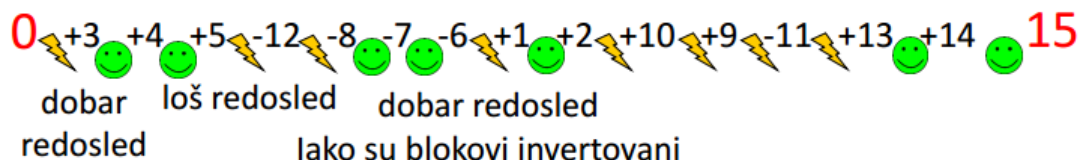
(-6 +1 +2 +3 +4 +5)
 (-5 -4 -3 -2 -1 +6)
 (+1 +2 +3 +4 +5 +6)

Slika 1.8: Kraći način

1.3 Teorema o prekidnoj tački

Uočimo da su uzastopni elementi (npr. (+12 +13)) poželjni, jer se javljaju u istom redosledu kao i u identičnoj permutaciji. Takodje, i (-11 -10) su poželjni, jer se mogu inverzijom postaviti u pravi redosled. Ova dva para elemenata imaju zajedničku osobinu da je drugi element za 1 veći od prvog. Stoga, definišemo pojmove suseda i prekida.

Definicija 1.4. Uređeni parovi (p_i, p_{i+1}) u permutaciji $P = (p_1, p_2, \dots, p_n)$ predstavljaju **susede**, ako je $p_{i+1} - p_i = 1$, a u suprotnom čine **prekid** (slika 1.9).



Slika 1.9: Susedi i prekidi

Primetimo da su permutaciji P na slici 1.9 dodate vrednosti 0 i $|P| + 1$. S obzirom da je bilo koji par uzastopnih elemenata permutacije predstavlja ili susede ili prekide, važi sledeća jednakost za proizvoljnu permutaciju P :

$$\#susedi(P) + \#prekidi(P) = |P| + 1.$$

U permutaciji P dužine n maksimalan broj suseda $n + 1$ i to važi samo za jediničnu permutaciju. Kod permutacije $(-n(-n - 1) \dots - 2 - 1)$ postoje dva prekida $((0, -1)$ i $(-1(n + 1)))$.

Na slici 1.10 prikazan je jedan scenario za sortiranje po promenama. Za svaku permutaciju prikazan je broj prekidnih tačaka (prekida) a deo koji će biti obrnut u narednom koraku je prikazan crvenom bojom. Broj prekidnih tačaka se nakon svakog koraka smanjuje i na kraju sortiranja se svodi na nulu pa sortiranje po promenama možemo smatrati eliminacijom prekidnih tačaka.

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

Slika 1.10: Sortiranje po promenama eliminacijom prekidnih tačaka

Razmotrimo koliki je maksimalan 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 (označeno crvenom) imamo pet prekidnih tačaka:

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

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

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

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.

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.

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)

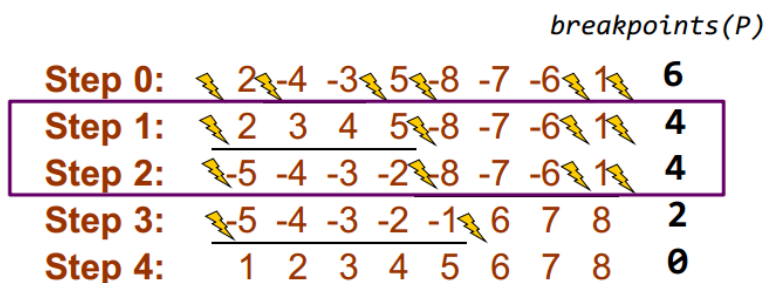
Granični prekidi su (+2 -4) i (-3 +5), nakon promene postanu (+2 +3) i (+4 +5) i oba para predstavljaju susede. Time je smanjen broj prekidnih tačaka unutar cele permutacije za dva.

Kao što je bilo reči, 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 nije manje od polovine broja prekidnih tačaka.

$$d_{rev}(P) \geq \#prekidi(P)/2$$

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 (slika 1.11, između koraka 1 i 2).



Slika 1.11: Sortiranje po promenama eliminacijom prekidnih tačaka

Teorema o prekidnoj tački određuje minimalni broj promena pri transformaciji permutacije P na jediničnu permutaciju. Razmotrimo maksimalan broj promena. Permutacija $(+n + (n-1) \dots +1)$

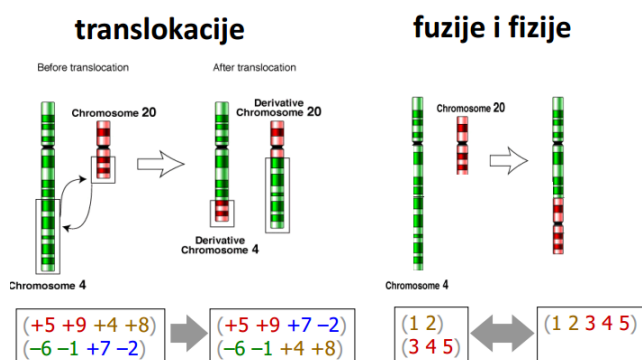
ima maksimalan broj prekidnih tačaka $(n+1)$ i za njeno sortiranje je potrebno $n+1$ promena. Prema teoremi o prekidnoj tački, data donja granica $((n+1)/2)$ je značajno manja od stvarne vrednosti što znači da to nije dobra aproksimacija rastojanja između proizvoljne permutacije P i jedinične permutacije. U nastavku ćemo predložiti drugačije strategije za aproksimiranje ovog rastojanja.

1.4 Preuređivanje u multihromozomalnim genomima

Umesto što posmatramo preuređivanje gena u okviru jednog hromozoma (hromozom X kod čoveka i miša), generalizujemo problem i posmatramo sve hromozome genoma. U multihromozomalnim genomima, pored obrtanja, postoje i drugi načini za preuređivanje genoma: translokacije, fuzije i fizije.

1.4.1 Translokacije, fuzije i fizije

Za modeliranje translokacija posmatramo multihromozomalni genom sa k hromozoma kao permutaciju koja je podeljena na k delova. Na primer, genom $(+1 +2 +3 +4 +5 +6) (+7 +8 +9 +10 +11)$ je sastavljen od dva hromozoma $(+1 +2 +3 +4 +5 +6)$ i $(+7 +8 +9 +10 +11)$.



Slika 1.12: Translokacije, fuzije i fizije

Translokacija podrazumeva razmenu segmenata različitih hromozoma. Na primer, translokacija dva hromozoma $(+1 +2 +3 +4 +5 +6) (+7 +8 +9 +10 +11)$ može dovesti do stvaranja sledeća 2 hromozoma: $(+1 +2 +3 +4 +9 +10 +11) (+7 +8 +5 +6)$. Možemo zamišljati translokaciju kao prvo cepanje svakog od hromozoma $(+1 +2 +3 +4 +5 +6) (+7 +8 +9 +10 +11)$ na 2 dela, $(+1 +2 +3 +4) (+5 +6) (+7 +8) (+9 +10 +11)$, a zatim lepljenje rezultujućih segmenata u 2 nova hromozoma, $(+1 +2 +3 +4 +9 +10 +11) (+7 +8 +5 +6)$.

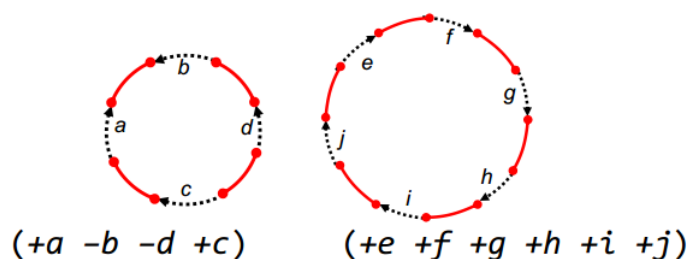
Preuređenja u multihromozomalnim genomima nisu ograničena na promene i translokacije. Ona takođe uključuju hromozomske fuzije, koje spajaju 2 hromozoma u 1, kao i fizije, koje dele 1 hromozom na 2 hromozoma (slika 1.12). Na primer, 2 hromozoma $(+1 +2 +3 +4 +5 +6) (+7 +8 +9 +10 +11)$ mogu biti fuzionisani (spojeni) u 1 hromozom $(+1 +2 +3 +4 +5 +6 +7 +8 +9 +10 +11)$. Sledeća fizija ovog hromozoma može dovesti do 2 hromozoma $(+1 +2 +3 +4) (+5 +6 +7 +8 +9 +10 +11)$.

Pre pet miliona godina, ubrzo nakon evolutivnog razdvajanja čoveka i šimpanze, fuzija dva hromozoma (nazvana 2A i 2B) u jednom od naših predaka stvorila je ljudski hromozom 2 i smanjila broj hromozoma sa 24 na 23.

1.5 Problem rastojanja 2-prekida

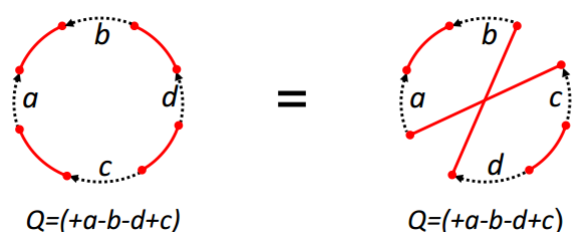
1.5.1 Od linearnih do cirkularnih hromozoma

Možemo pretpostaviti da su svi hromozomi u genomu cirkularni. Iako ova pretpostavka ne odgovara biološkoj realnosti, pojednostaviće dalju analizu bez uticaja na konačni zaključak. Fokusaćemo se na jedan od hromozoma u multihromozomalnom genomu i razmotriti transformacije promene kružnog hromozoma $P = (+ a -b -c + d)$ u $Q = (+ a -b -d + c)$. Kružne hromozome ćemo predstaviti grafovski pomoću dve vrste grana: crnih usmerenih i crvenih neusmerenih (slika 1.13). Cne usmerene grane predstavljaju blokove sintenije. Crvene neusmerene grane povezuju susedne blokove sintenije.



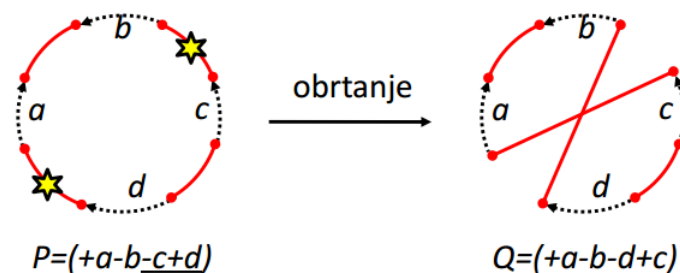
Slika 1.13: Hromozomi P i Q

Jedan hromozom ima više grafovskih reprezentacija u zavisnosti od toga kojim redom navodimo blokove sintenije odnosno kojim redom u grafu navedemo postavimo crne grane. Slika 1.14 prikazuje dve takve ekvivalentne reprezentacije za hromozom Q.



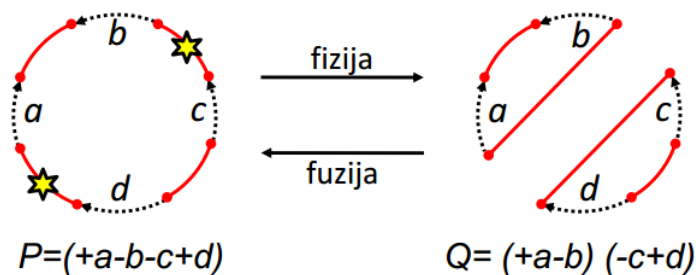
Slika 1.14: Ekvivalentne reprezentacije hromozoma Q

Slika 1.15 ilustruje promenu kojom se hromozom P transformiše u hromozom Q obrtanjem susednih blokova $-c +d$. Grane u grafu za hromozom Q su predstavljene u istom redosledu kao za hromozom P pa stoga uočavamo da se obrtanje na grafu predstavlja ukrštanjem njegovih grana koje nisu prethodno bile ukrštene, ili obrnuto. Kao što možemo videti, promena briše dve crvene grane iz P (povezivanje b sa c i d sa a) i zamenjuje ih sa dve nove crvene grane (povezivanje b sa d i c sa a).



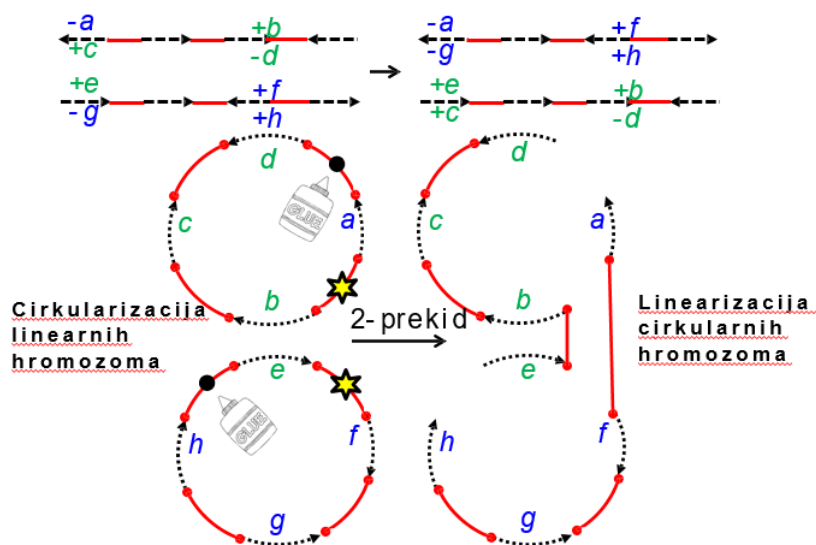
Slika 1.15: Obrtnanje

Slika 1.16 ilustruje fiziju $P = (+ a -b -c + d)$ u $Q = (+ a -b) (- c + d)$. Inverzna operacija fiziji odgovara fuziji dva hromozoma iz Q u hromozom P . Operacije fuzije i fizije, kao i promene, odgovaraju brisanju dve grane u jednom genomu i njihovim zamenjivanjem sa 2 nove grane u drugom genomu.



Slika 1.16: Fizija i fuzija

Translokacija koja uključuje dva linearna hromozoma (kao što je prikazano na slici 1.12) takođe se može simulirati tako što hromozome najpre cirkularizujemo, a zatim dve crvene grane zamenimo sa dve različite crvene grane, kao što je prikazano na slici 1.17.



Slika 1.17: Translokacija

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

Definicija 1.5. Rastojanje 2-prekida $d(P, Q)$ predstavlja minimalni broj 2-prekida koji transformišu genom P u genom Q.

Definicija 1.6. Problem rastojanja 2-prekida: Naći rastojanje 2-prekida između dva genoma.

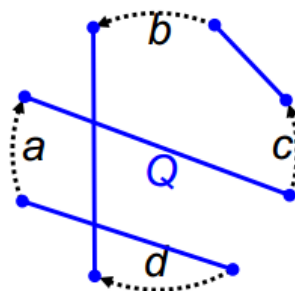
Ulaz. Dva genoma nad istim skupom blokova sintenije.

Izlaz. Rastojanje 2-prekida između ovih genoma.

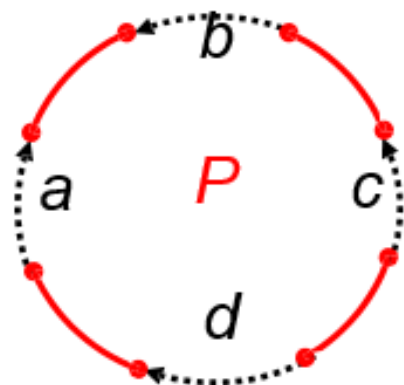
1.6 Problem rastojanja 2-prekida

Za izračunavanje rastojanja 2-prekida između hromozoma P i Q predstavimo oba genoma grafovski na slikama 1.18 i 1.19.

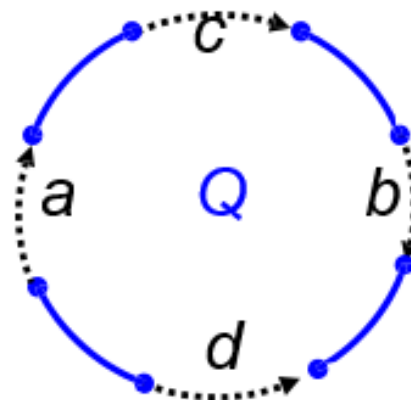
Genom Q mozemo predstaviti i na drugi način (slika 1.20).



Slika 1.20: Drugaćija reprezentacija genoma Q

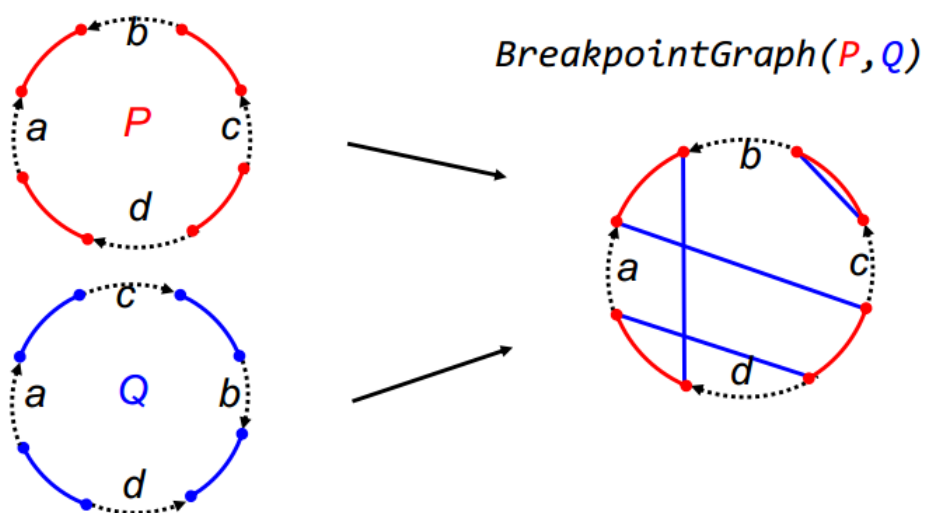


Slika 1.18: Genom P



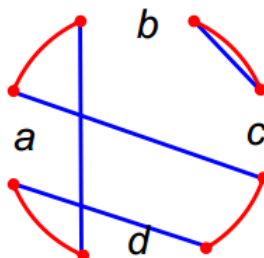
Slika 1.19: Genom Q

Preklapanjem grafova genoma P i Q dobijamo njihovu zajedničku reprezentaciju, takozvani *prekidni graf* u oznaci $BreakpointGraph(P, Q)$ (slika 1.21).



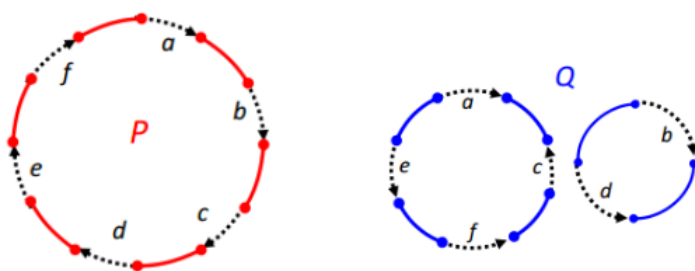
Slika 1.21: Preklapanje grafova genoma P i Q

Crvene i crne grane u prekidnom grafu formiraju genom P. **Plave** i crne grane u prekidnom grafu formiraju genom Q. **Crvene** i **plave** grane formiraju **alternirajuće crveno-plave cikluse** (slika 1.22). Označimo sa $cycle(P, Q)$ broj alternirajućih crveno-plavih ciklusa.

$BreakpointGraph(P, Q)$ 

Slika 1.22: Alternirajući crveno-plavi ciklusi

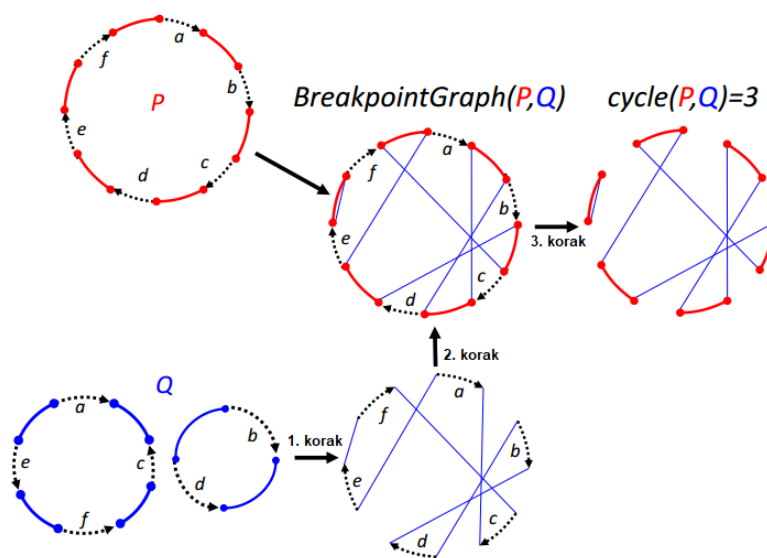
Kako bismo analizirali šta predstavlja broj alternirajućih crveno-plavih ciklusa $cycle(P, Q)$, posmatrajmo grafove genoma P i Q na slici 1.23.



Slika 1.23: Grafovi genoma P i Q

Kreiranje prekidnog grafa za genome P i Q obuhvata sledeće korake prikazane na slici 1.24:

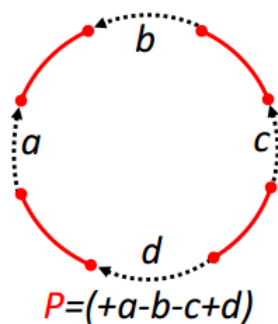
1. korak: Promena redosleda crnih grana u reprezentaciji genoma Q tako da odgovara redosledu crnih grana u genomu P
2. korak: Preklapanje grafova genoma P i Q u jedan
3. korak: Uklanjanje crnih grana



Slika 1.24: $cycle(P,Q)$

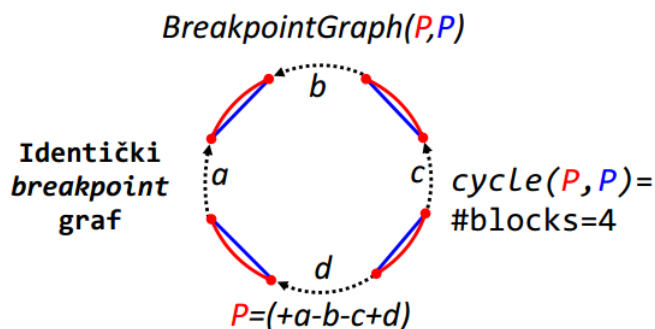
Ukupan broj alternirajućih crveno-plavih ciklusa iznosi 3: jedan ciklus je određen blokovima e i f, jedan blokovima d, c, a i b (redom), i jedan preostalim blokovima.

Posmatrajmo sada jednostavniji genom P (slika 1.25) i razmotrimo koje Q za dato P maksimizuje $cycle(P, Q)$.



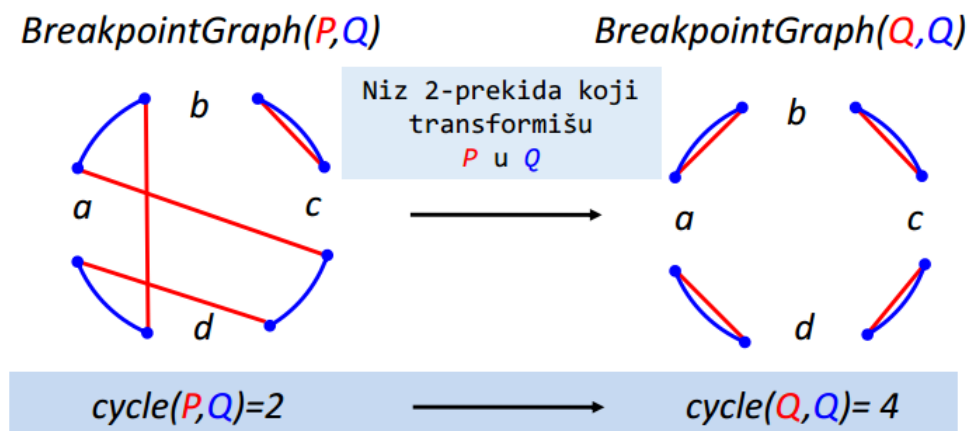
Slika 1.25: Genom P

Pretpostavimo da P i Q imaju isti broj blokova sintenije i označimo taj broj sa $blocks(P, Q)$. Ako su P i Q identični, njihov prekidni graf se sastoji od $blocks(P, Q)$ ciklusa dužine 2 od kojih svaki sadrži jednu crvenu i jednu plavu granu. Cikluse dužine 2 nazivamo **identičkim ciklusima**, a prekidni graf koji se sastoji od ovih ciklusa nazivamo **identičkim prekidnim grafom** (slika 1.26).



Slika 1.26: Identički prekidni graf

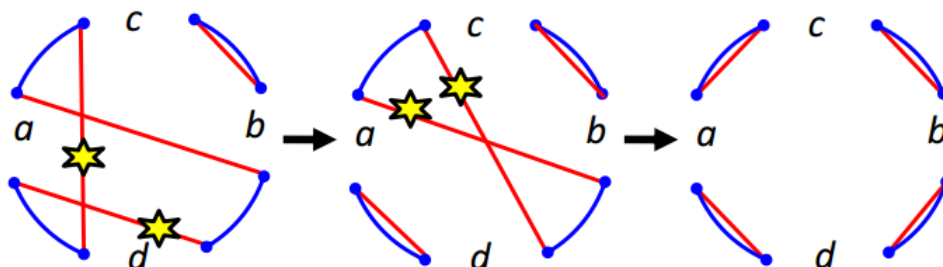
Analizirajmo kako preuređenje genoma utiče na crveno-plave cikluse. Svaka transformacija $P \rightarrow Q$ odgovara transformaciji između dva prekidna grafa kojom jedan prekidni graf postaje identički (slika 1.27).



Slika 1.27: Transformacija $P \rightarrow Q$

Preuređenja genoma koja P transformišu u Q takođe utiču i na $cycle(P, Q)$ (slika 1.28). Sa svakim preuređenjem, $cycle(P, Q)$ se povećava i na kraju transformacije dostiže maksimum jednak $cycle(Q, Q)$ koji je jednak $blocks(P, Q)$ broju blokova sintenije od kojih se sastoje P i Q . Možemo zaključiti da se tokom transformacije $P \rightarrow Q$ broj alternirajućih crveno-plavih ciklusa poveća za $blocks(P, Q) - cycle(P, Q)$.

$$\begin{aligned}
 P &= (+a -b -c +d) \rightarrow P' = (+a -b -c -d) \rightarrow P'' = Q = (+a +c +b -d) \\
 \text{BreakpointGraph}(P, Q) &\rightarrow \text{BreakpointGraph}(P', Q) \rightarrow \text{BreakpointGraph}(Q, Q) \\
 \text{cycle}(P, Q) &= 2 \rightarrow \text{cycle}(P', Q) = 3 \rightarrow \text{cycle}(Q, Q) = 4
 \end{aligned}$$



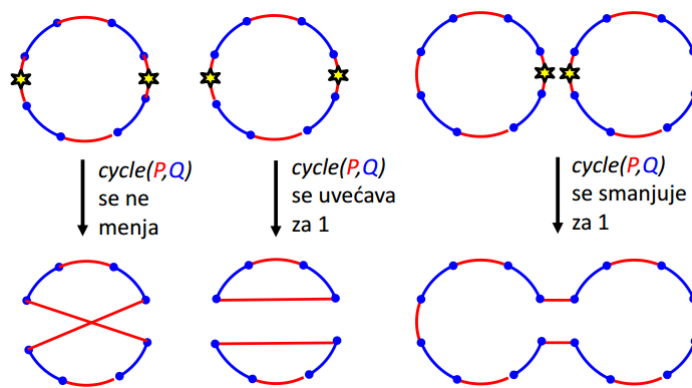
Slika 1.28: Uticaj preuređenja na cycle(P,Q)

U prethodnom primeru (slika 1.28) videli smo da se obrtanjem $\text{cycle}(P, Q)$ može uvećati za 1. Razmotrimo koliko druge vrste 2-prekida doprinose uvećanju broja ciklusa.

Ciklusna teorema 2-prekid može povećati $\text{cycle}(P, Q)$ za najviše 1.

Dokaz Posmatrajmo sliku 1.29. 2-prekid može ukloniti dve crvene grane koje ili pripadaju istom ciklusu ili različitim ciklusima. U prvom slučaju, 2-prekid ili ne menja $\text{cycle}(P, Q)$ (levo) ili uvećava za 1 (sredina), a u drugom slučaju (desno) umanjuje za 1.

Formalno, 2-prekid dodaje dve nove crvene grane kojima formira najviše dva nova ciklusa. U isto vreme, 2-prekid i briše dve crvene grane kojima uklanja najmanje 1 stari ciklus. Stoga, broj crveno-plavih ciklusa je uvećan za najviše $2 - 1 = 1$. \square



Slika 1.29

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 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 potvrđan odgovor na ovo pitanje.

Teorema o rastojanju 2-prekida. Rastojanje 2-prekida između genoma P i Q je jednako $blocks(P, Q) - cycle(P, Q)$.

Dokaz. Svaka transformacija genoma P u genom Q, koja se sastoji od niza pojedinačnih promena, povećava broj alternirajućih ciklusa za $blocks(P, Q) - cycle(P, Q)$. Rastojanje 2-prekida $d(P, Q)$ predstavlja minimalni broj 2-prekida koji genom P transformišu u genom Q, pa možemo reći da je $d(P, Q) \geq blocks(P, Q) - cycle(P, Q)$. 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 dve crvene i dve plave grane. Svaki neidentički ciklus se, kao što je prikazano na slici 1.29 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, što je i trebalo dokazati. \square

Rastojanje 2-prekida između genoma čoveka i miša

- Eksperimentalno je pokazano da se genomi čoveka i miša se mogu rastaviti na 280 blokova sintenije (dužine bar pola miliona nukleotida)
- Graf prekidnih tačaka nad ovim blokovima ima ukupno 35 ciklusa
- Na osnovu teoreme o rastojanju 2-prekida:

$$d(H, M) = blocks(H, M) - cycle(H, M) = 280 - 35 = 245$$

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

Literatura