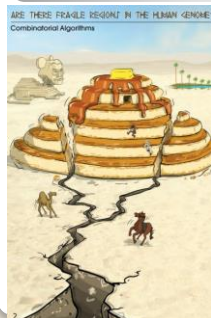


# G: Przetaszowania w genomie. Algorytmy kombinatoryczne, algorytmy zachłanne



230

## Problem przetasowań w genomie

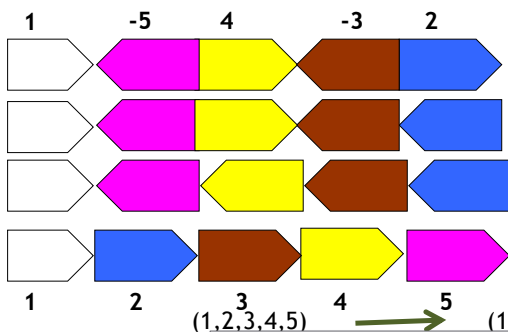


Mitochondrialne genotypy kapusty i rzepy są w 99% identyczne

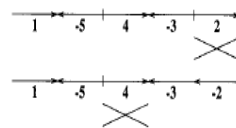
- Jeffrey Palmer (1980)



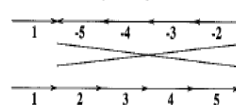
Różne jest uporządkowanie genów w mitochondrialnym DNA



*B. oleracea*  
(cabbage)



*B. campestris*  
(turnip)

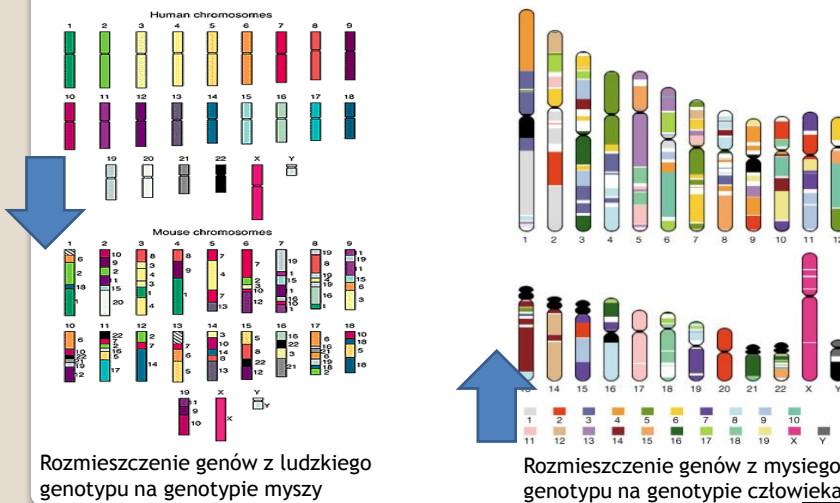


3 przetasowania, to jest 3 permutacje ze znakiem

### Od myszy do człowieka

Zestawy genów są IDENTYCZNE (prawie) ale są one różnie poustawiane.

Uważa się, że poprzez zmianę porządku genów ujawnia się EWOLUCJA.



Danuta Makowiec, Algorytmika dla bioinformatyki, kurs 2018/2019

### Przetasowanie w genomie: odwrócenia

**Odwrócenie:**  
najprostsza postać przetasowania

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

oznaczenie:

$\rho(4,8)$

skąd

dokąd

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

Odwrócenie kolejności w spójnej grupie wraz odwróceniem kierunku odczytu.

Danuta Makowiec, Algorytmika dla bioinformatyki, kurs 2018/2019

## Przetaszowania w genomie: od myszy do człowieka

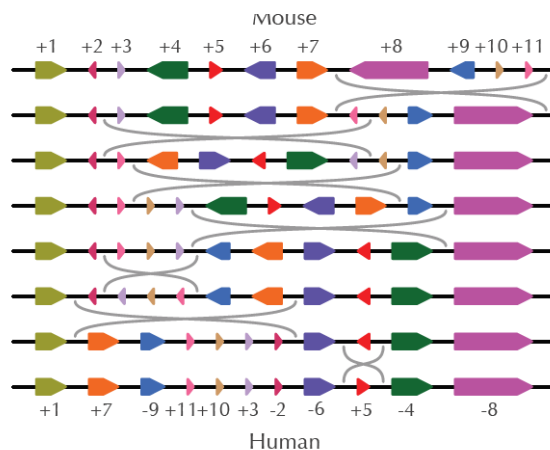


FIGURE 6.6 Encoding the mouse X chromosome as the identity permutation implies encoding the human X chromosome as (+1 +7 -9 +11 +10 +3 -2 -6 +5 -4 -8).

## Przetaszowania w genomie

- Niech geny w genomie są oznaczane kolejnymi liczbami naturalnymi: 1,2, ...,n przy czym kierunek genu jest oznaczany '+', '-' lub ''
- Przetaszowaniem nazywamy każdą permutację ze znakiem  $\pi$  tego ciągu.
 
$$\pi = (\pi_1 \pi_2 \dots \pi_n)$$
- Ilość przetaszowań transformujących daną permutację  $\pi$  w drugą  $\pi^*$  nazywać będziemy odległością  $d(\pi, id)$  w odwróceniach.
- W przypadku gdy  $\pi^* = id$  to taką odległość oznaczamy  $d_{rev}(\pi) = d(\pi, id)$

Przykład:

$$\begin{aligned} \rho(2,3) \pi_0 &: (+2 \text{ -4 } \text{-3} +5 \text{ -8 } \text{-7 } \text{-6 } +1) \\ \rho(5,7) \pi_1 &: (+2 +3 +4 +5 \text{ -8 } \text{-7 } \text{-6 } +1) \\ \rho(8,8) \pi_2 &: (+2 +3 +4 +5 +6 +7 +8 \text{ +1}) \\ \rho(1,7) \pi_3 &: (\text{+2 } \text{+3 } \text{+4 } \text{+5 } \text{+6 } \text{+7 } \text{+8 } \text{-1}) \\ \pi_4 &: (\text{-8 } \text{-7 } \text{-6 } \text{-5 } \text{-4 } \text{-3 } \text{-2 } \text{-1}) \\ \rho(1,8) \pi_5 &: (+1 +2 +3 +4 +5 +6 +7 +8) \end{aligned}$$

W pięciu przetaszowaniach uporządkowaliśmy bloki genomu

### Algorytm zachłanny sortowania przez przetasowania

Przykład: poszerzanie prefiksowe

```
(+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)
```

- Mówimy, że element  $k$  permutacji  $\pi$  jest **wysortowany** jeżeli  $\pi_k = k$
- Mówimy, że permutacja  $\pi$  jest  **$k$ -wysortowana** jeśli elementy  $1, \dots, k$  są wysortowane
- Istnieje dokładnie jedna operacja odwrócenia sortująca  $k$ -ty element, zachowująca porządek pierwszych  $(k-1)$  elementów.

**GREEDYSORTING( $P$ )**

```
approxReversalDistance  $\leftarrow$  0
for  $k \leftarrow 1$  to  $|P|$ 
  if element  $k$  is not sorted
    apply the  $k$ -sorting reversal to  $P$ 
    approxReversalDistance  $\leftarrow$  approxReversalDistance + 1
  if the  $k$ -th element of  $P$  is  $-k$ 
    apply the  $k$ -sorting reversal to  $P$ 
    approxReversalDistance  $\leftarrow$  approxReversalDistance + 1
return approxReversalDistance
```

Czy jest optymalny?

### Punkty łamiące permutacji

**Definicja:** Parę sąsiadujących elementów  $\pi_i \pi_{i+1}$  w permutacji  $\pi$  nazywamy **uzgodnioną** jeżeli  $\pi_{i+1} - \pi_i = 1$ .

W przeciwnym wypadku parę  $\pi_i \pi_{i+1}$  nazywamy **punktem łamiącym**

Przykład:

↑ +3 +4 +5 -12 -8 -7 -6 +1 +2 +10 +9 -11 +13

8 punktów łamiących

Niech:  $\text{Adj}(\pi)$  to ilość par uzgodnionych w  $\pi$   
 $\text{BreakPoints}(\pi)$  to ilość punktów łamiących w  $\pi$

Zauważ, że:

$\text{Adj}(\pi) + \text{BreakPoints}(\pi) = n+1$  dla  $\pi$  o  $n$  elementach  
 $\text{Adj}(\pi = \text{id}) = n+1$

## Sortowanie przez odwracanie obniżające ilość punktów łamiących

Przykład:

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

FIGURE 6.7 A sorting by reversals. The inverted interval of each reversal is shown in red, while breakpoints in each permutation are marked by vertical blue segments.

Twierdzenie:

$d_{\text{rev}}(\pi)$  jest większe lub równe  $\text{BreakPoints}(\pi)/2$

## Przetaskowania multichromosomalne w genomie

Translokacje, fuzje i rozszczepienia w genomie

Translokacja to wymiana segmentów pomiędzy chromosomami

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

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

Można pomyśleć, że ta translokacja powstała wskutek:

Rozszczepienia chromosomów:

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

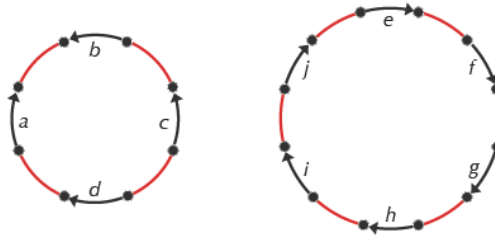
I następnie przetaskowanego połączenia

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

## Graf genomu

241

Przyjmijmy, że chromosomy są kołowe



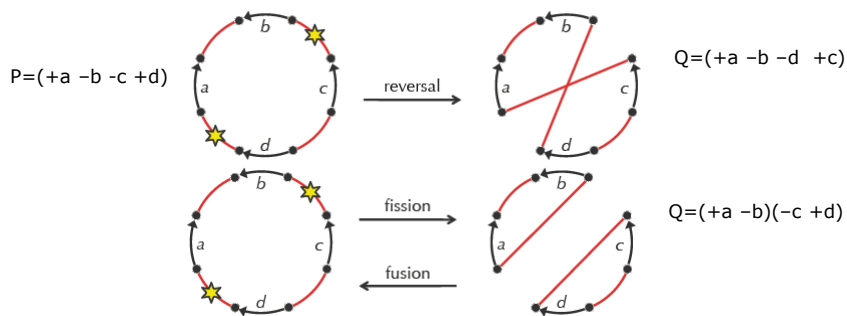
**FIGURE 6.9** A genome with two circular chromosomes,  $(+a -b -c +d)$  and  $(+e +f +g +h +i +j)$ . Black directed edges represent syntenic blocks, and red undirected edges connect adjacent syntenic blocks. A circular chromosome with  $n$  elements can be written in  $2n$  different ways; the chromosome on the left can be written as  $(+a -b -c +d)$ ,  $(-b -c +d +a)$ ,  $(-c +d +a -b)$ ,  $(+d +a -b -c)$ ,  $(-a -d +c +b)$ ,  $(-d +c +b -a)$ ,  $(+c +b -a -d)$ , and  $(+b -a -d +c)$ .

Danuta Makowiec, *Algorytmika dla bioinformatyki*, kurs 2018/2019

2019-01-23

## Translokacje, fuzje i rozszczepienia w genomie

242



Operacje na grafie genomu polegającą na zamianie dwóch nieskierowanych krawędzi na dwie inne, ale pomiędzy tymi samymi węzłami grafu nazywamy **2-break**

### 2-Break Distance Problem:

Find the 2-break distance between two genomes.

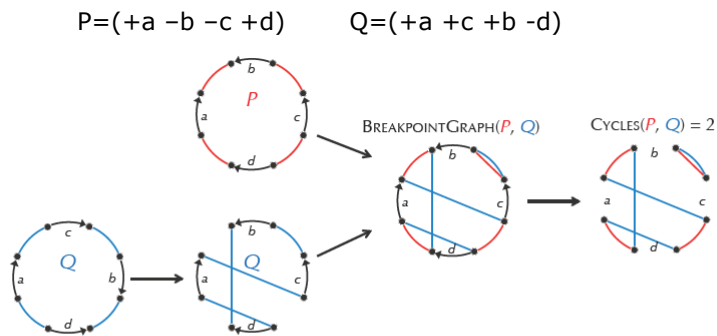
**Input:** Two genomes with circular chromosomes on the same syntenic blocks.

**Output:** The 2-break distance between these genomes.

Danuta Makowiec, *Algorytmika dla bioinformatyki*, kurs 2018/2019

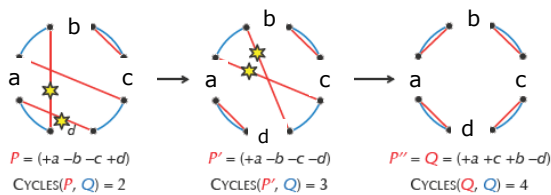
2019-01-23

**Grafy punktów łamiących**



**FIGURE 6.14** Constructing the breakpoint graph for the unichromosomal genomes  $P = (+a -b -c +d)$  and  $Q = (+a +c +b -d)$ . After rearranging the black edges of  $Q$  so that they are arranged the same as in  $P$ , the breakpoint graph  $BREAKPOINTGRAPH(P, Q)$  is formed by superimposing the graphs of  $P$  and  $Q$ . As shown on the right, there are two alternating red-blue cycles in this breakpoint graph.

**Sortowanie poprzez 2-break punkty łamiące**



**FIGURE 6.18** Every 2-break transformation of genome  $P$  into  $Q$  corresponds to a transformation of  $BREAKPOINTGRAPH(P, Q)$  into  $BREAKPOINTGRAPH(Q, Q)$ . In the example shown, the number of red-blue cycles in the graph increases from  $CYCLES(P, Q) = 2$  to  $CYCLES(Q, Q) = BLOCKS(Q, Q) = 4$ .

**2-Break Sorting Problem:**

Find a shortest transformation of one genome into another by 2-breaks.

**Input:** Two genomes with circular chromosomes on the same synteny blocks.

**Output:** The sequence of genomes resulting from applying a shortest sequence of 2-breaks transforming one genome into the other.