

Algorithms in Genome Research

Winter 2015/2016

Exercises

Number 10, Discussion: 2016 January 29

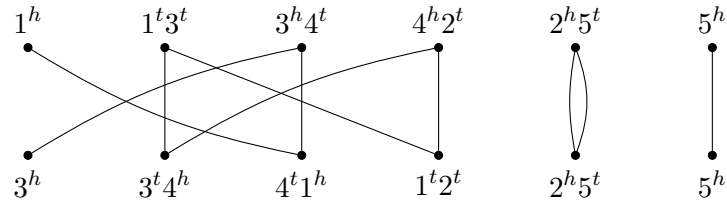
1. **Reversal distance.** What is the reversal distance for the following permutations?
 - (a) $\pi = (0 \ -3 \ 1 \ 2 \ 4 \ 6 \ 5 \ 7 \ -15 \ -13 \ -14 \ -12 \ -10 \ -11 \ -9 \ 8 \ 16)$
 - (b) $\pi = (0 \ 3 \ 10 \ 9 \ 4 \ 2 \ 1 \ 5 \ 7 \ 6 \ 8 \ 11 \ 13 \ 12 \ 14)$.
 - (c) $\pi = (0 \ 1 \ 3 \ -5 \ 4 \ 6 \ 2 \ 7)$
 - (d) $\pi = (0 \ 2 \ 4 \ 3 \ 5 \ 1 \ 6 \ 8 \ 7 \ 9)$
2. Given a genome A with l_1 linear chromosomes, and B with l_2 linear chromosomes, how many paths does the multichromosomal breakpoing graph $BP(A, B)$ have?
3. **DCJ distance.** Given the gene set $\mathcal{G} = \{1, 2, 3, 4, 5, 6\}$ and genomes $A = \{1^t, 4^h 6^h, 3^h 1^h, 4^t 2^t, 5^h 6^t, 2^h, 3^t, 5^t\}$ and $B = \{1^t, 1^h 2^t, 2^h 3^t, 3^h 4^t, 4^h 5^t, 5^h 6^t, 6^h\}$, answer the following questions:
 - (a) Draw genomes A and B .
 - (b) Draw the BP graph $BP(A, B)$. What is the DCJ distance between A and B ?
 - (c) Find a sequence of DCJ operations transforming A into B .
4. Consider the following algorithm for sorting by DCJ:

Algorithm 2 (Greedy sorting by DCJ)

```
1: for each adjacency  $\{p, q\}$  in genome  $B$  do
2:   let  $u$  be the element of genome  $A$  that contains  $p$ 
3:   let  $v$  be the element of genome  $A$  that contains  $q$ 
4:   if  $u \neq v$  then
5:     replace  $u$  and  $v$  in  $A$  by  $\{p, q\}$  and  $(u \setminus \{p\}) \cup (v \setminus \{q\})$ 
6:   end if
7: end for
8: for each telomere  $\{p\}$  in genome  $B$  do
9:   let  $u$  be the element of genome  $A$  that contains  $p$ 
10:  if  $u$  is an adjacency then
11:    replace  $u$  in  $A$  by  $\{p\}$  and  $(u \setminus \{p\})$ 
12:  end if
13: end for
```

- (a) Prove that in each iteration, either the number of cycles of BP increases by one, or the number of odd paths increases by 2. *Tip: draw the adjacency pq of B in the BP, and what types of elements u and v that genome A can have. There are 3 possibilities, depending if u and v are adjacencies or telomeres. What changes in the BP after the DCJ is applied?*

5. **The Adjacency Graph.** Another useful graph for comparing genomes is the *adjacency graph*. In this graph, the vertices are the adjacencies and telomeres of A and B , and there is an edge between vertices if they have extremities in common. For instance, for genomes $A = \{1^h, 1^t 3^t, 3^h 4^t, 4^h 2^t, 2^h 5^t, 5^h\}$ and $B = \{3^h, 3^t 4^h, 4^t 1^h, 1^t 2^t, 2^h 5^t, 5^h\}$ the adjacency graph is



- (a) Draw the adjacency graph of genomes $A = \{1^t, 4^h 6^h, 3^h 1^h, 4^t 2^t, 5^h 6^t, 2^h, 3^t, 5^t\}$ and $B = \{1^t, 1^h 2^t, 2^h 3^t, 3^h 4^t, 4^h 5^t, 5^h 6^t, 6^h\}$ (same as exercise 3).
- (b) What is the difference between the BP graph and the adjacency graph?